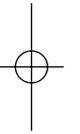
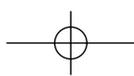
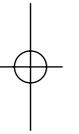
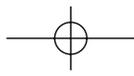
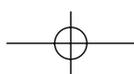


THINGS



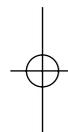




# Things

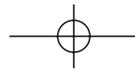
*Papers on Objects, Events, and Properties*

STEPHEN YABLO



**OXFORD**  
UNIVERSITY PRESS





**OXFORD**

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.  
It furthers the University's objective of excellence in research, scholarship,  
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi  
Kuala Lumpur Madrid Melbourne Mexico City Nairobi  
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece  
Guatemala Hungary Italy Japan Poland Portugal Singapore  
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press  
in the UK and in certain other countries

Published in the United States  
by Oxford University Press Inc., New York

© in this volume Stephen Yablo 2010

The moral rights of the author have been asserted  
Database right Oxford University Press (maker)

First published 2010

All rights reserved. No part of this publication may be reproduced,  
stored in a retrieval system, or transmitted, in any form or by any means,  
without the prior permission in writing of Oxford University Press,  
or as expressly permitted by law, or under terms agreed with the appropriate  
reprographics rights organization. Enquiries concerning reproduction  
outside the scope of the above should be sent to the Rights Department,  
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover  
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data  
Data available

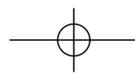
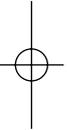
Library of Congress Cataloging in Publication Data  
Library of Congress Control Number: 2010927223

Typeset by Laserwords Private Limited, Chennai, India  
Printed in Great Britain  
on acid-free paper by

MPG Books Group, Bodmin and King's Lynn

ISBN 978-0-19-926648-7 (Hbk.)  
ISBN 978-0-19-926649-4 (Pbk.)

1 3 5 7 9 10 8 6 4 2



## Acknowledgements

I thank the editors and publishers who have granted permission to reprint the papers appearing in this volume. Dates and first places of publication are as follows:

'Identity, Essence, and Indiscernibility': *Journal of Philosophy* 84/6 (1987), 293–314

'Intrinsicness': *Philosophical Topics* 26 (1999), 479–505

'Cause and Essence': *Synthese* 93 (1992), 403–49

'Advertisement for <sup>a</sup>Sketch of an Outline of a Prototheory of Causation': *Causation and Counterfactuals*, edited by John Collins, Ned Hall, and L. A. Paul (MIT Press, 2004), 119–37

'Does Ontology Rest on a Mistake?': *Supplement to Proceedings of the Aristotelian Society* 72/1 (1998), 229–62

'Apriority and Existence': *New Essays on the A Priori*, edited by Paul Boghossian and Christopher Peacocke (Oxford University Press, 2000), 197–228

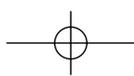
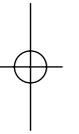
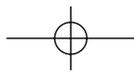
'Go Figure': *Midwest Studies in Philosophy* 25 (2001), 72–93

'Abstract Objects': *Philosophical Issues* 12 (2002), 220–40

'The Myth of the Seven': *Fictionalism in Metaphysics*, edited by Mark Eli Kalderon (Oxford University Press, 2005), 88–115

'Non-Catastrophic Presupposition Failure': *Content and Modality*, edited by Judith Thomson and Alex Byrne (Oxford University Press, 2006), 164–90

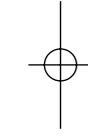
'Must Existence-Questions Have Answers?' *Metametaphysics*, edited by David Chalmers, David Manley, and Ryan Wasserman (Oxford University Press, 2009), 507–25

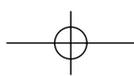
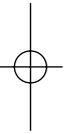
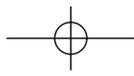




## *Contents*

Introduction	1
1. Identity, Essence, and Indiscernibility	13
2. Intrinsicness	33
3. Cause and Essence	59
4. Advertisement for a Sketch of an Outline of a Prototheory of Causation	98
5. Does Ontology Rest on a Mistake?	117
6. Apriority and Existence	145
7. Go Figure: A Path through Fictionalism	177
8. Abstract Objects: A Case Study	200
9. The Myth of the Seven	221
10. Carving Content at the Joints	246
11. Non-Catastrophic Presupposition Failure	269
12. Must Existence-Questions Have Answers?	296
<i>Index</i>	315



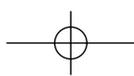
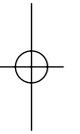
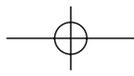




The fundamental-seeming philosophical question, How much of our science is merely contributed by language and how much is a genuine reflection of reality? is perhaps a spurious question which itself arises wholly from a certain particular type of language.

Quine, 'Identity, Ostension, and Hypostasis'







## Introduction

This book is not exactly a continuation of my PhD thesis. But the two have a lot in common.

1. Both are called *Things*.
2. They share an epigraph.
3. Both contain lots of unashamed first-order metaphysics.
4. The first-order topics are similar: identity, essence, properties, and causation.<sup>1</sup>
5. Both question whether anything is really at stake in certain first-order debates.
6. The questioning connects up, in both cases, with first-order philosophy of a non-metaphysical kind—philosophy of language and mathematics, mainly.
7. Both owe their existence, ultimately, to the author of the shared epigraph.



The last point requires some explanation. If you learn about essentialism from Quine, as I did, then you are going to find it outrageous, as I also did. Anti-essentialism was the closest thing I had to a religious identity in my first year of college. *Naming and Necessity* shook my complacency, but I couldn't bring myself to switch sides. (I became a lapsed anti-essentialist, continuing to worship at the site of my old abandoned church.) I wrestled with these issues for a long time, convinced that Quine and Kripke had to be each somehow right about something.

“Identity, Essence, and Indiscernibility” was my attempt at a resolution. I tried to design a world that essentialists and anti-essentialists could both feel at home in. It's a world with two identity-like relations: identity proper, and coincidence, a demodalized analogue requiring sharing of “categorical” properties only—properties with nothing hypothetical about them.<sup>2</sup> Anti-essentialists need coincidence because it plays, for them, the role of identity. Essentialists need it too, though, if they are not to deny the obvious: that things whose differences are all on the counterfactual side are still in some sense the same.

<sup>1</sup> Papers on the metaphysics of mind appeared in an earlier volume, *Thoughts* (Oxford, 2009).

<sup>2</sup> Kripke speaks of the “dark doctrine of a relation of ‘contingent identity.’” Coincidence is meant to be the salvageable core of that doctrine.



Suppose a bolt snaps suddenly. How does its *suddenly* snapping relate to its simply snapping? Davidson sees no difference between them. Any difference would have to be on the score of suddenness, he seems to be thinking; and *both* events are sudden.

But we talk as if there are two events here, not one. You might ask, for instance, whether the bridge would still have collapsed, had the snapping not been sudden. But there is no sensible question of what would have happened had the bolt's *suddenly* snapping not been sudden.<sup>3</sup> The question makes no sense because it is only the snapping that persists into worlds where the bolt snaps gradually; its suddenly snapping is not so modally flexible.

"The bridge disaster was caused by the bolt's suddenly snapping, not by its snapping as such." Davidson finds this hard to make sense of.<sup>4</sup> How can the one event be more efficacious than the other, when the other was every bit as sudden? The answer is that only one of the pair was *essentially* sudden, and differences of this kind—"merely modal" differences—bring a lot in their wake. They bear on a thing's transworld career, hence on its counterfactual behavior, hence on its causal relations. It was the bolt's suddenly snapping that caused the disaster, because the bridge would not have collapsed, had the snapping taken more time.

What if the bridge would have collapsed either way? Then, I claim, the cause was the bolt's (simply) snapping. For causes are expected to be proportional to their effects; they should include a good bit of what the effect requires, and not too much that was not required. Proportionality is judged by asking

- (i) whether *e* would still have occurred, had *c* occurred in the absence of richer alternatives (like the bolt's suddenly snapping), and
- (ii) whether it would still have occurred had poorer alternatives (like the bolt's failing somehow or other) occurred without *c* occurring.

"Cause and Essence" develops this idea at some length, applying it eventually to the epiphenomenalism debate. (There is more about epiphenomenalism in "Mental Causation" and "Wide Causation," which appeared in *Thoughts*.)

The proportionality condition assumes, however, that effects counterfactually depend on their causes in the first place. And in some cases they appear not to. The window would still have broken even if Suzy hadn't thrown her rock, because Billy's rock was coming along just behind. I attempt to restore the lost dependence by holding fixed those aspects of the situation whereby other would-be causes are frozen out of the action. Billy's rock never touches the glass, after all. Holding that fixed, the window would *not* have broken, if Suzy

<sup>3</sup> This is not just because of the de dicto impossibility of a "sudden" snapping being in the same world gradual. The de dicto impossibility also obtains when we ask what the bridge would have done had the *sudden* snapping been gradual. And there is no absurdity about this at all. The bolt's sudden snapping is accidentally sudden, just as the dog's empty bowl is only accidentally empty.

<sup>4</sup> Davidson (1967).  $\wedge$

("Sudden" is, I assume, a relative term.)

hadn't thrown. This much was already in the dissertation. "Advertisement for a Sketch of an Outline of a Proto-Theory of Causation" attempts a specification of what can be held fixed. The effect depends on  $c$  modulo some choices of  $G$ , and  $c$  is otiose modulo other choices—it is over and above events that would have done the job all by themselves.  $C$  causes  $e$ , the claim is, if  $e$  depends on  $c$  modulo some relatively natural  $G$ —one more natural than any  $G$  modulo which  $c$  is otiose.<sup>5</sup>

The properties that coincident entities share—categorical properties—are modal analogues of the properties that Lewisian duplicates share—intrinsic properties. Categorical properties are to logical space, more or less, as intrinsic properties are to physical space. This suggests a possible analysis: a property is categorical if its distribution in world  $w$  is an intrinsic feature of  $w$ ; what goes on in other worlds is irrelevant. The problem with this analysis is that the notion of an intrinsic property is no clearer than that of a categorical property.

But the analogy still has something to teach us. The categorical properties are characterized in "Identity, Essence, and Indiscernibility" as the properties that  $x$  and  $x^+$  are bound to agree on, if  $x^+$  is a refinement of  $x$ . Refinement is a modal counterpart of the relation wholes bear to their parts—so perhaps the intrinsic properties are definable in terms of part/whole. This is what I attempt in "Intrinsicness." A property  $P$  is intrinsic iff a thing in world  $w$  cannot gain or lose  $P$  in the transition to  $w^+$ , where  $w$  is part of  $w^+$ .<sup>6</sup>

A word now about the metametaphysical papers.<sup>7</sup> These papers have sometimes been read as nominalist screeds, but that was never the intent. The intent was to ask in a hermeneutic spirit what Phyllis the mathematical physicist is talking about, when she says, *Star formation is an exponentially decreasing function of time elapsed since redshift 2*.<sup>8</sup> She is not talking (I claim) about the function whose value on input  $m$  is the number of newly formed stars in the  $m$ th millennium after redshift 2.<sup>9</sup> She is talking about stars, and how they used to pop up more often than they do now.

Is it possible to say in general what Phyllis is asserting when she utters a math-infused sentence  $S$ ? Probably not; but her message is *on the whole* better captured by  $S$ 's concrete content  $\|S\|$  than its literal content  $|S|$ —where

$\|S\|$  is the proposition true in a world  $w$  iff  $S$  is true in some  $v$  concretely indiscernible from  $w$ , albeit perhaps richer than  $w$  in mathematical objects.<sup>10</sup>

<sup>5</sup> For discussion see Björnsson (2007), Longworth, MS, and Hall and Paul (forthcoming).

<sup>6</sup> Parsons develops an interesting objection to this approach, to which Clark (forthcoming) responds.

<sup>7</sup> This work is discussed in Burgess (2004), Colyvan (forthcoming), Eklund (2005), Gallois (1998), Linnebo (MS and forthcoming), Manley (2009), Rayo (2008), Rosen and Burgess (2005), and Stanley (2001).

<sup>8</sup> Redshift 2 light dates back to two or three billion years after the Big Bang.

<sup>9</sup> " $y$  is such and such a function of  $x$ " might just mean " $y$  varies in such and such a way with  $x$ ." But functions certainly do turn up sometimes in mathematical physics; assume for example's sake that it happens here.

<sup>10</sup> I first encountered this sort of definition in Gideon Rosen's dissertation.

Hermeneutic fictionalists think that

(HF) Phyllis asserts only the concrete content  $||S||$  of  $S$ .

Why is this confused with (hermeneutic) nominalism? Because concrete content is *nominalistic* content; and because the *nominalist* maintains (HF) as well. The reason it's a confusion is that the hermeneutic nominalist adds a second thesis, perhaps as an explanation of *why* no more is asserted:

(HN) It is only the concrete content  $||S||$  that is true.

Why *else* would Phyllis hesitate to assert the full content  $|S|$ , if not for the reason just given: the full content might not be true?

To answer a question with a question: the “why else?” gets its rhetorical force from the assumption that

(\*) One ought, other things equal, to assert as much of a sentence's full content as (one thinks) might be true

—and why should we accept that assumption? There are two serious problems with it. First, “*might be true*” sets the bar too low. If Phyllis has no idea whether the platonistic bits are true, that gives her all the reason she needs not to assert them. Second, one is not even expected to assert as much of  $|S|$  as one *knows* to be true. There is, if anything, an opposite expectation. One should *presuppose* as much as possible; otherwise assertive content becomes a portmanteau any part of which could be the point really at issue.<sup>11</sup>

An admittedly fanciful analogy may help to clarify the nature of Phyllis's speech act, according to the hermeneutic fictionalist. Imagine that someone (Harry, call him) goes apoplectic listening to Bill O'Reilly. *Wow, that guy really gets his goat*, you say. Obviously you are not talking about anyone's actual goat when you say this. Why not, though? Are you holding back because you think your remark would be false, construed as a remark about goats? Not at all; Harry may have a goat, for all you know. Is it that you are not sure that your remark would be, on a literal construal, true? Not that, either. You are not talking about Harry's goat even if you know for a fact that he has one. The remark is just orthogonal to any views you might have about goats.

The same goes for Phyllis's remark about star formation being an exponentially decreasing function of time passed since redshift 2. She may in her private moments believe in functions. The remark may be literally true, in her view. She may even take satisfaction in the fact that  $S$ 's full content is something that she believes. But Phyllis is not, in uttering  $S$ , putting that proposition forward as true. She is registering an opinion about stars, not functions.

<sup>11</sup> Stalnaker (1974). A different reason, emphasized by linguists, to beef up presuppositional content is that one otherwise suggests that the foregone additions might be untrue. See Heim (1991), Sauerland (2004), and Schlenker, MS on the principle *Maximize Presupposition*.

How is it determined whether a given utterance of  $S$  conveys its nominalistic content  $||S||$  or its full, platonistic, content  $|S|$ ? One place to look is conversational uptake. If Phyllis's students cannot claim to know, on the strength of her testimony, that there are exponentially decreasing functions, this suggests that nothing has been asserted about them. (Audiences normally come away knowing what an expert speaker has asserted to them.) That her students would not feel in the least misled, should Phyllis turn out to be a nominalist, suggests her views about this played no role in what she was saying. (A speaker who asserts what she takes to be false is open to charges of dishonesty.)

The question so far has been, How do we determine that  $S$  is not being used to assert its full literal content  $|S|$ ? Another question is, What is the rule that determines the content that *is* being asserted? There are various options here.

*Meta-Fictionalism*

$S$ 's assertive content is that  $S$  is true according to the Story of Standard Mathematics, or properly pretendable in the Standard Mathematical Game. This is a non-starter, I think. Phyllis is no more talking about stories and games than about functions.

*Object Fictionalism* (or a descendant of it called *Figuralism*)

$S$ 's assertive content is the real-world fact that *makes*  $S$  true in the story, or pretendable in the game. This casts  $||S||$  in the role of  $S$ 's *metaphorical* content, if we understand metaphors as moves in prop-oriented make believe games.<sup>12</sup>

*Presuppositionalism*

$S$ 's assertive content is the "logical remainder" when the content  $|\pi|$  of operative presuppositions is subtracted from  $|S|$ . If we define  $|S| - |\pi|$  as the part of  $|S|$  that is not about whether  $|\pi|$ , this makes  $||S||$  the part of  $|S|$  that is not about whether mathematical objects exist.

*Subject-Matter-ism*

$S$ 's assertive content is the part of  $|S|$  that IS about the subject matter under discussion. If we are doing physics, this makes  $||S||$  the part of  $|S|$  that concerns the physical world.

Several of the papers here defend Object Fictionalism. Presuppositionalism begins to make an appearance in "Non-Catastrophic Presupposition Failure" and "Must Existence-Questions Have Answers?" Subject matters will be bearing some of the representational load in future work.

The four -isms are different, but they have one thing in common: they leave the question of abstract objects' existence wide open. How is that question to be resolved? Quine's ontological program is no help here, I argue in "Does

<sup>12</sup> Walton (1993). One example he gives is *Crotone is in the arch of the Italian boot*. This is used to assert the fact about Crotone's real-world location that licenses us in pretending that it's in the arch of the Italian boot.

Ontology Rest on a Mistake?” The paper *suggests* that questions like this may be unanswerable; there may be, as the expression goes, no fact of the matter either way. It makes no attempt, however, to explain how such a thing is possible.

“Non-Catastrophic Presupposition Failure” offers the beginnings of a model. By a typical numerical statement, let’s mean the kind encountered in the marketplace, not the philosophy room. Let  $\pi$  be what a typical numerical statement presupposes, and let  $a$  be what it asserts. Our first two assumptions are

1. Numbers are typically presupposed; they figure in  $\pi$ , not  $a$ .
2. The presupposition is *fail-safe*— $a$  evaluates the same whether  $\pi$  is true or not.

Assumption 3 derives from Frege’s Context Principle, which tells us to seek the meanings of words in their sentence-level effects. (The principle holds of abstract terms if it holds anywhere; it is not clear where else their meaning is to be sought.)

3. Whether numerals refer is determined by their effects on truth-value.

I assume, finally, that the question of numbers’ existence is not in practice distinguishable from the question of whether numerals refer. The idea that although 7 exists, it is not the referent of ‘7’, or that ‘7’ refers, but to something other than 7, is self-evidently absurd.

4. Numbers exist if and only if numerals refer.

The argument now proceeds as follows:

5. Numerals’ effect on truth-value is the same whether they refer or not. (by 2)
6. There is nothing to determine whether numerals refer. (by 3 and 5)
7. There is no fact of the matter about whether numerals refer. (by 6)
8. There is no fact of the matter as to whether numbers exist. (by 4 and 7)

The model is horribly crude. But let me try to explain the underlying idea. It goes back to something Russell says in *Introduction to Mathematical Philosophy*.

When you have taken full account [of our notion of Napoleon, still] you have not touched the actual man; but in the case of Hamlet, you have come to the end of him. If no one thought about Hamlet, there would be nothing left of him; if no one had thought about Napoleon, he would have soon seen to it that some one did.

This is a suggestive passage which does not, however, wear its meaning on its face. What it suggests to me is an ontological decision procedure—a way of deciding of some putative object X whether it exists (see Figure 0.1). Napoleon exists, because there is more to him than can be recovered from our ideas about what Napoleon is supposed to be like. He has the feature, let us say, of *surpassing expectations*. It is because Napoleon surpasses expectations that the assumption

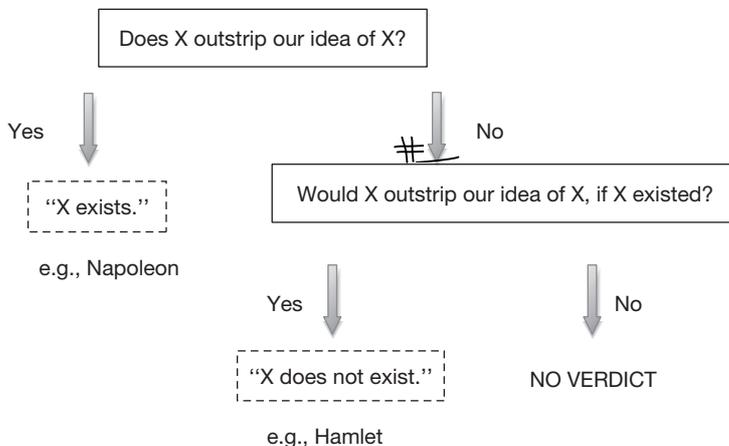


Figure 0.1

bottom of 'g' seems greyed out

of his existence is not fail-safe. Napoleon the existing man is an original source of information of the type that decides truth-values. To look at it from the other end, his name's sentence-level effects are due in part to the path Napoleon wound up taking through history, *given that he was here*. If we imagine him out of history, he takes the sentence-level effects with him.

Hamlet would share this feature—the feature of *surpassing expectations*—if he existed.<sup>13</sup> He would be a person, and people have blood types, first loves, private misgivings, and so on—none of which can be made up in their absence on the basis of casting notes. If, as Russell seems to suggest, Hamlet does *not* surpass expectations, then we may infer from the fact that he *would* surpass them, if he existed, that Hamlet does not exist.

Now we come to things which neither surpass expectations, nor would surpass them, if they were real. I call things like this *preconceived* because they are constrained not to get too far out in front of how we think of them. Either they *should* have feature F, given their job description, or they *don't* have feature F.<sup>14</sup>

A thing is preconceived if its principal features are fixed by its job description. This does not mean one can easily see *how* they are fixed. Our logical powers are limited, and the type of fixing at issue may not even be codifiable—as, for example, the second-order consequence relations that determine arithmetical truth are not codifiable. Also, some preconceived entities may have their properties fixed not absolutely, but modulo the features of certain other, non-preconceived, entities on which they constitutively depend. ({Socrates} is preconceived relative to Socrates.) These are crucial caveats which will be taken as understood.

<sup>13</sup> It might be enough that a thing has the *potential* to surpass expectations. I will ignore this subtlety.

<sup>14</sup> Benacerraf (1965).

if possible

A useful comparison here is with Voltaire's God: the God we would have had to invent, had he not existed. I take it that the need to invent him—to stipulate his existence—arises because there are things an existent God can do that the mere job description cannot. (Saving us comes to mind.) Preconceived objects are different in this respect. They make the same contribution uninvented as invented. Anything they themselves might have accomplished, truth-value-determination-wise, is accomplished already by their job description.

How are we to tell whether things like this exist? Premise 3—*Whether numerals refer is determined by their effects on truth-value*—suggests it's not going to be easy; one can make sense of the truth-value effects on either hypothesis. It might seem that only God can know what is really responsible for the observed distribution of Trues and Falses. But this rests on a misunderstanding of premise 3. It is not about how language-users “determine” (in an epistemological sense) whether numerals refer; it's about how the world determines whether numerals refer. Even God, then, must judge abstract existence from behind a veil of truth-value-determination. There is no possibility even for him of a “language-unblinkered inspection of the contents of the world, of which the outcome might be to reveal that there was indeed nothing there capable of serving as the referents of . . . numerical singular terms.”<sup>15</sup>

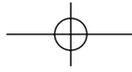
I have been asked (by Gideon Rosen) how my dismissive attitude toward abstract ontology is to be squared with the exorbitant concrete ontology of “Identity, Essence, and Indiscernibility.” That paper puts a plenitude of enduring things at the site of each concrete object, one for each conceivable modal career running through the object's this-worldly manifestation.

Here is my best guess about what is going on. It's the ontology of *preconceived* objects that I have trouble with; and the realm of the preconceived extends beyond pure abstracta. It extends to impure abstracta like {Socrates}; they, as already mentioned, are preconceived relative to the source of their impurity. The realm of the preconceived extends, perhaps, to some purely concrete objects as well. The mereological sum of my eyes and Obama's ears is, so it seems, preconceived relative to my eyes and his ears. To come back around finally to the essentialist multitudes of “Identity, Essence, and Indiscernibility,” they are preconceived relative to the concrete objects of which they represent alternative modal inflections. I didn't worry about appealing to them in the definition of categorical property for the same sort of reason as I don't worry about quantifying over numbers today.

I do worry about something else. Objects are preconceived if they are kept on a short leash; they are not allowed to run too far out ahead of our basic assumptions about them. This *prima facie* conflicts, however, with another cherished theme, the “open texture of concepts.”<sup>16</sup> Numbers could in principle surprise us. They could fall subject to outside pressures, exerted, for instance, by the larger structures in which we want to embed them. The definition we all

<sup>15</sup> Wright (1983: 13–14).

<sup>16</sup> See various papers in *Thoughts*.



learned of “prime,” for instance, doesn’t extend well to the complex plane.<sup>17</sup> (A prime number is not supposed to have proper factors. But 5 is the product of  $2+i$  and  $2-i$ .) The proper definition, apparently, is based on what, in elementary number theory, is treated as a theorem:  $p$  is prime if to divide the product of  $a$  and  $b$ , it must divide  $a$  or  $b$ . Now, if a known customer like *primeness* has in fact to do, not with a number’s factors, but its multiples, who knows what other surprises might be in store?<sup>18</sup> Intuitions about probability have been brought to bear against proposed axioms of set theory. Mereological sums could turn out to follow other physical laws than one would suppose from the laws governing their parts. The point is that a short leash can always be lengthened. This is not a contradiction, I think, not any more than fallibilism about analytic truth is a contradiction. But it bothers me all the same.

The papers overlap somewhat; pages 229 ff, for instance, repeat material from pages 158 ff, which material also occurs on pages 129 ff, where footnote 43 describes it as borrowed from a still earlier paper. (I tell you now so you won’t be mad later.)

Who to thank? For advice and ideas—Louise Antony, Kent Bach, Karen Bennett, Anne Bezuidenhout, Paul Boghossian, John Burgess, David Chalmers, Mark Crimmins, Cian Dorr, Andy Egan, Matti Eklund, Hartry Field, Kit Fine, Kai von Fintel, Graeme Forbes, Danny Fox, Bas van Fraassen, André Gallois, Tamar Gendler, Mario Gómez-Torrente, Bob Hale, Ned Hall, Sally Haslanger, Allen Hazen, John Hawthorne, Irene Heim, David Hills, Eli Hirsch, Thomas Hofweber, Richard Holton, Lloyd Humberstone, David Kaplan, Jaegwon Kim, Saul Kripke, Rae Langton, David Lewis, Francis Longworth, Penelope Maddy, Vann McGee, Ruth Millikan, Sarah Moss, Laurie Paul, Peter Railton, Agustín Rayo, Gideon Rosen, Carolina Sartorio, Jonathan Schaffer, Stephen Schiffer, Laura Schroeter, Sydney Shoemaker, Alan Sidelle, Ted Sider, Scott Soames, Bob Stalnaker, Jason Stanley, Zoltán Szabó, Jamie Tappenden, Judy Thomson, Gabriel Uzquiano, David Velleman, Ken Walton, Ralph Wedgwood, Tim Williamson, and Crispin Wright. For seeing the book into print—Abigail Coulson, Kate Williams, and Peter Momtchiloff. Students, thanks for listening. MIT, thanks for being there. Linguists, thanks for the explanations. Yablangers, thanks for the answer to “what is there?”: everything.

## REFERENCES

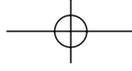
- Benacerraf, Paul (1965). “What numbers could not be”, *Philosophical Review* 74/1 47–73.
- Björnsson, Gunnar (2007). “How Effects Depend on Their Causes, Why Causal Transitivity Fails, and Why We Care about Causation”, *Philosophical Studies* 133/3, 349–90.

<sup>17</sup> Thanks here to Jamie Tappenden.

<sup>18</sup> See Tappenden (1995 and 2008), Wilson (2006), and Pincock, MS.

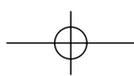
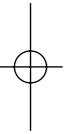
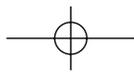


- Burgess, John (2004). "Mathematics and bleak house", *Philosophia Mathematica* 12/1, 18–36.
- Clark, Michael (forthcoming). "Inclusionism and the Problem of Unmarried Husbands." To appear in *Erkenntnis*.
- Colyvan, Mark (forthcoming). "There is No Easy Road to Nominalism." To appear in *MIND*.
- Davidson, Donald (1967). "Causal Relations", *Journal of Philosophy* 64, 691–703.
- Eklund, Matti (2005). "Fiction, Indifference, and Ontology", *Philosophy and Phenomenological Research* 71/3, 557–579.
- Gallois, André (1998). "Does Ontology Rest on a Mistake?", *Aristotelian Society Supplementary Volume* 72, 263–283.
- Hall, Ned, and Paul, Laurie (forthcoming). *Causation and its Counterexamples: A Traveler's Guide*.
- Heim, Irene (1991). 'Artikel und definitheit', in A. von Stechow and D. Wunderlich (eds.), *Semantics: An International Handbook of Contemporary Research* (Berlin: Walter de Greyter).
- Kripke, A. Saul (1980). *Naming and Necessity* (Cambridge, Mass: Harvard University Press).
- Linnebo, Øystein, MS. "Ontology and the Concept of an Object".  
 — (forthcoming). "Pluralities and Sets", to appear in *Journal of Philosophy*.
- Longworth, Francis, MS. "Causation is Not De Facto Dependence".
- Mancosu, Paolo (ed.) (2008). *The Philosophy of Mathematical Practice* (New York: Oxford University Press).
- Manley, David (2009). "When Best Theories Go Bad", *Philosophy and Phenomenological Research* 78/2, 392–405.
- Munitz, Milton, and Unger, Peter (eds.) (1974). *Semantics and Philosophy* (New York: New York University Press).
- Parsons, Josh (2007). "Is everything a world?" *Philosophical Studies* 134/2, 165–181.
- Pincock, Chris, MS. Critical Notice of Wilson (2006).
- Rayo, Agustin (2008). "On specifying truth-conditions", *The Philosophical Review* 117/3, 385–443.
- Rosen, Gideon, and Burgess, John (2005). "Nominalism Reconsidered", in Stewart Shapiro (ed.) *The Oxford Handbook of Philosophy of Mathematics and Logic* (Oxford: OUP), 515–535.
- Sauerland, Uli (2004). "On embedded implicatures", *Journal of Cognitive Science* 5/1, 107–37.
- Schlenker, Philippe, MS. "Maximize Presupposition and Gricean Reasoning".
- Stalnaker, Robert (1974). "Pragmatic Presuppositions", in Munitz and Unger (eds.) (1974), repr. in Stalnaker (1999).  
 — (1999). *Context and Content* (Oxford: Oxford University Press).
- Stanley, Jason (2001). "Hermeneutic Fictionalism," *Midwest Studies in Philosophy* 25, 36–71.
- Tappenden, Jamie (2008). "Mathematical Concepts and Definitions", in Mancosu (2008), pp. 256–75.



- Tappenden, Jamie (1995). "Extending Knowledge and 'Fruitful Concepts': Fregean Themes in the Foundations of Mathematics", *Noûs* 29/4, 427–67.
- Walton, Kendall (1993). 'Metaphor and Prop Oriented Make-Believe', *European Journal of Philosophy* 1/1, 39–57.
- Wilson, Mark (2006). *Wandering Significance: An Essay on Conceptual Behaviour* (Oxford: Oxford University Press).
- Wright, Crispin (1983). *Frege's Conception of Numbers as Object* (Aberdeen: Aberdeen University Press).



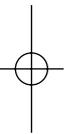




# 1

## Identity, Essence, and Indiscernibility

Can things be identical *as a matter of fact* without being *necessarily* identical? Until recently it seemed they could, but now “the dark doctrine of a relation of ‘contingent identity’”<sup>1</sup> has fallen into disrepute. In fact, the doctrine is worse than disreputable. By most current reckonings, it is refutable. That is, philosophers have *discovered* that things can never be contingently identical. Appearances to the contrary, once thought plentiful and decisive, are blamed on the befuddling influence of a powerful alliance of philosophical errors. How has this come about? Most of the credit goes to a simple argument (due to Ruth Marcus and Saul Kripke) purporting to *show* that things can never be only contingently identical. Suppose that  $a$  and  $\beta$  are identical. Then they share all their properties. Since one of  $\beta$ 's properties is that it is necessarily identical with  $\beta$ , this must be one of  $a$ 's properties too. So necessarily  $a$  is identical with  $\beta$ , and it follows that  $a$  and  $\beta$  cannot have been only contingently identical.<sup>2</sup>



Donald Davidson, Sally Haslanger, Kit Fine, David Kaplan, Noa Latham, Shaughan Lavine, Barry Loewer, George Myro, Sydney Shoemaker, Robert Stalnaker, and David Velleman all made comments that helped me with the writing of this paper.

<sup>1</sup> Saul Kripke, *Naming and Necessity* (Cambridge, Mass.: Harvard, 1980), p. 4.

<sup>2</sup> In the prevailing necessitarian euphoria, it has become difficult to recapture the atmosphere of a few years back, when contingent identity was a commonplace of logical and metaphysical theorizing. To cite just two examples, Dana Scott's "Advice on Modal Logic" [in Karel Lambert, ed., *Philosophical Problems in Logic* (Boston: D. Reidel, 1970)] urged that "two individuals that are generally distinct might share all the same properties (of a certain kind!) with respect to the present world . . . Hence they are equivalent or *incident* at the moment. Relative to other points of reference they may cease to be incident" (165). And most of the early mind/body-identity theorists—U. T. Place, J. J. C. Smart, Thomas Nagel, among others—took themselves to be asserting the contingent identity of mental and physical entities. Smart, for instance, says very explicitly that "on the brain-process thesis the identity between the brain process and the experience is a contingent one" ["Sensations and Brain Processes," *Philosophical Review*, LXVIII, 2 (April 1959):141–156, p. 152]. There is a question, actually, how it is that so many people *thought* that an impossible thing was possible. One hypothesis—maybe it is Kripke's hypothesis—is that these people were just very mixed up. And, in fact, it does seem that to varying degrees they were. Ruth Marcus once described W. V. Quine as thinking that modal logic was conceived in sin, the sin of confusing use and mention. Contingent identity had, if anything, even shadier beginnings, because at least three separate sins attended at its conception. Contingency was routinely identified with (or at least thought to follow from) *a posteriority*; particular identity statements (like "this pain is identical to that brain-event") were insufficiently distinguished from general identity statements (like "consciousness is a process in the brain"); and the contingent truth of an identity statement was equated with the contingency of the asserted identity, guaranteeing that contingent coincidence of concepts would be taken for

## I. A PARADOX OF ESSENTIALISM

Despite the argument's simplicity and apparent cogency, somehow, as Kripke observes, "its conclusion . . . has often been regarded as highly paradoxical."<sup>3</sup> No doubt there are a number of bad reasons for this (Kripke himself has exposed several), but there is also a good one: essentialism without some form of contingent identity is an untenable doctrine, because essentialism has a shortcoming that only some form of contingent identity can rectify. The purpose of this paper is to explain, first, why contingent identity is required by essentialism and, second, how contingent identity is permitted by essentialism.

Essentialism's problem is simple. Identicals are indiscernible, and so discernibles are distinct. Thus, if  $\alpha$  has a property necessarily which  $\beta$  has only accidentally, then  $\alpha$  is distinct from  $\beta$ . In the usual example, there is a bust of Aristotle, and it is formed of a certain hunk of wax. (Assume for the sake of argument that the hunk of wax composes the bust throughout their common duration, so that temporal differences are not in question.) If the bust of Aristotle is *necessarily* a bust of Aristotle and if the hunk of wax is only *accidentally* a bust of Aristotle, then the bust and the hunk of wax are not the same thing. Or suppose that Jones drives home at high speed. Assuming that her speeding home is something *essentially* done at high speed, whereas her driving home only *happens* to be done at high speed, her speeding home and her driving home are distinct.

So far, so good, maybe; but it would be incredible to call the bust and the wax, or the driving home and the speeding home, distinct, and leave the matter there. In the first place, that would be to leave relations between the bust and the hunk of wax on a par with either's relations to the common run of *other* things, for example, the Treaty of Versailles. Secondly, so far it seems an extraordinarily baffling metaphysical coincidence that bust and wax, though entirely distinct, nevertheless manage to be *exactly alike* in almost every ordinary respect: size,

the contingent identity of the things specified. So evidence for the confusion hypothesis is not lacking. The other hypothesis is, of course, that people recognized, confusedly perhaps, something sensible and defensible in the notion that things can be identical as a matter of fact. Philosophically, it does not much matter which of these hypotheses is correct. There *is* something sensible and defensible in the idea of contingent identity, whether its advocates recognized it or not. Or so I hope to show.

<sup>3</sup> Kripke, "Identity and Necessity," in Stephen P. Schwartz, ed., *Naming, Necessity, and Natural Kinds* (Ithaca, N.Y.: Cornell, 1977), p. 67. Let me say at the outset that, as far as I can see, Kripke's argument does succeed in establishing what it claims to establish, namely that identity, in the strict sense, can never obtain contingently. If the conclusion seems paradoxical, as it surely does, that is because people are confusing it with the genuinely paradoxical thesis that there can be *no* relation with the characteristics traditionally associated with "contingent identity." Speaking more generally, what Kripke says about identity is important and correct, and not questioned here; but people may have thought that his conclusions closed off certain avenues of investigation which are in fact still open. And that is perhaps why some of those conclusions have seemed hard to accept.

weight, color, shape, location, smell, taste, and so on indefinitely. If distinct statues (say) were as similar as this, we would be shocked and amazed, not to say incredulous. How is such a coincidence possible? And, thirdly, if the bust and the wax are distinct (pure and simple), how is it that the number of middle-sized objects on the marble base is not (purely and simply) 2 (or more)? Ultimately, though, none of these arguments is really needed: that the bust and the wax are in *some* sense the same thing is perfectly obvious.

Thus, if essentialism is to be at all plausible, nonidentity had better be compatible with intimate identity-like connections. But these connections threaten to be inexplicable on essentialist principles, and essentialists have so far done nothing to address the threat.<sup>4</sup> Not *quite* nothing, actually; for essentialists have tried to understand certain (special) of these connections in a number of (special) ways. Thus, it has been proposed that the hunk of wax *composes* the bust; that the driving home *generates* the speeding home; that a neural event *subserves* the corresponding pain; that a computer's structural state *instantiates* its computational state; that humankind *comprises* personkind; and that a society is *nothing over and above* its members. Now all these are important relations, and each is importantly different from the others. But it is impossible to ignore the fact that they seem to reflect something quite general, something not adequately illuminated by the enumeration of its special cases, namely, the phenomenon of things' being distinct *by nature* but the same *in the circumstances*. And what is that if not the—arguably impossible—phenomenon of things' being contingently identical but not necessarily so? The point is that, if essentialism is true, then many things that are obviously in *some* sense the same will emerge as strictly distinct; so essentialism must at least provide for the possibility of intimate identity-like connections between distinct things; and such connections seem to be ways of being contingently identical. Essentialism, if it is to be plausible, has to be tempered by some variety of contingent identity.

<sup>4</sup> Observing that not only modal but temporal differences “establish that a statue is not the hunk of stone, or the congeries of molecules, of which it is composed,” Kripke allows that “mere non-identity . . . may be a weak conclusion” (“Identity and Necessity,” p. 101). *Extremely* weak, from the point of view of philosophical materialism. That pains were not identical with neural stimulations seemed to be a powerfully antimaterialistic result; but now it turns out to be compatible with pains and neural stimulations' being as tightly bound up with one another as statues and their clay. And what materialist would not be delighted with that result? On the other hand, “The Cartesian modal argument . . . surely can be deployed to maintain relevant stronger conclusions as well” (“Identity and Necessity,” p. 101). Possibly this means that the statue is “nothing over and above” its matter, whereas the same cannot be said of a person; in the sense that necessarily, the statue (but not the person) exists if its matter does, and with a certain organization (*Naming and Necessity*, p. 145). But it seems doubtful whether the statue *is* “nothing over and above” its matter *in that sense* (what if the statue's matter had gathered together by chance, before the earth was formed? what if a different sculptor had organized the matter?); and the subtler the sense in which a statue really is “nothing over and above” its matter, the less implausible it becomes that, in a substantially similar sense, a person is “nothing over and above” *its* matter. So there may still be room for doubt whether modal arguments establish significantly more difference between a person and her matter than between a statue and its.

Hence, essentialism is confronted with a kind of paradox: to be believable it needs contingent identity; yet its principles appear to entail that contingent identity is not possible. To resolve the paradox, we have to ask: What is the “nature” of a particular thing?

## II. ESSENCE

Begin with a particular thing  $a$ . How should  $a$  be characterized? That is, what style of characterization would best bring out “what  $a$  is”? Presumably a characterization of any sort will be via certain of  $a$ ’s properties. But which ones?

Why not begin with the set of all  $a$ ’s properties whatsoever, or what may be called the *complete profile* of  $a$ ? Since  $a$ ’s properties include, among others, that of being identical with  $a$ , there can be no question about the sufficiency of characterization by complete profile. But there may be doubt about its philosophical interest. For the properties of  $a$  will generally be of two kinds: those which  $a$  had to have and those which it merely happens to have. And, intuitively, the properties  $a$  merely happens to have reveal nothing of what  $a$  is, as contrasted with what it happens to be like. As Antoine Arnauld explains in a letter to Leibniz,

... it seems to me that I must consider as contained in the individual concept of myself only that which is such that I should no longer be me if it were not in me: and that all that is to the contrary such that it could be or not be in me without my ceasing to be me, cannot be considered as being contained in my individual concept.<sup>5</sup>

(Adding: “That is my idea, which I think conforms to everything which has ever been believed by all the philosophers in the world”!) If  $a$ ’s nonnecessary properties reveal nothing about what  $a$  is, nothing will be lost if they are struck from its characterization.

Dropping  $a$ ’s nonnecessary properties from its complete profile yields the set of all properties that  $a$  possesses essentially, or what can be called the *complete essence* of  $a$ .<sup>6</sup> Since  $a$  is essentially identical with  $a$ , the property of so being will be included in  $a$ ’s complete essence; so the sufficiency of the characterization is again beyond doubt. Nor can there be much question that complete essences do better than complete profiles at showing what particulars are by nature. But worries about philosophical interest remain.

In the first place, the essence of an entity ought, one feels, to be an assortment of properties *in virtue of which* it is the entity in question. But this requirement is

<sup>5</sup> H. T. Mason, ed., *The Leibniz–Arnauld Correspondence* (Manchester: University Press, 1967), p. 30.

<sup>6</sup> In some contexts it is useful to distinguish between essential and necessary and between accidental and contingent properties. For example, someone might think that, whereas Socrates is essentially human, he is only necessarily Greek-or-not. The distinction is intuitive but irrelevant to our purposes.

trivialized by the inclusion, in essences, of identity properties, like that of being identical with California. A thing does not get to be identical with California by having the property of so being, but gets to be California and to have that property, alike, by having certain *other* properties. And it is these other properties that really belong in a thing's characterization. Another way of putting what is probably the same point is that identity properties and their ilk are not "ground floor," but dependent or supervenient. As a kind of joke, someone I know explains the difference between his two twin collies like this: "It's simple: *this* one's *Lassie*, and *that* one's *Scottie*." What makes this a joke is that that cannot be all there is to it; and the reason is that identity properties are possessed not *simpliciter*, but dependently on other properties. It is only these latter properties that ought, really, to be employed in a thing's characterization.

Secondly, the essence of a thing is supposed to be a measure of *what is required* for it to be that thing. But, intuitively, requirements can be more or less. If the requirements for being  $\beta$  are stricter than the requirements for being  $\alpha$ , then  $\beta$  ought to have a "bigger" essence than  $\alpha$ . To be the Shroud of Turin, for instance, a thing has to have everything it takes to be the associated piece of cloth, *and* it has to have enshrouded Jesus Christ (this is assuming that the piece of cloth did, in fact, enshroud Jesus Christ). Thus, more is essential to the Shroud of Turin than to the piece of cloth, and the Shroud of Turin ought accordingly to have the bigger essence. So, if essences are to set out the requirements for being their possessors, it should be possible for one thing's essence to include another's.<sup>7</sup> What is perhaps surprising, however, is that this natural perspective on things will not survive the introduction of identity properties and their ilk into individual essences. Think of the piece of cloth that makes up the Shroud of Turin (call it "the Cloth of Turin"): if the property of being identical with the Cloth of Turin is in the Cloth of Turin's essence, then, since that property is

<sup>7</sup> Where *kinds* of things are concerned, this is comparatively uncontroversial: the essence attaching to the kind *cow* strictly includes the essence attaching to the kind *animal*. But, as Leibniz noticed, individuals can be thought of as instancing smallest or least kinds, what we might call *individual kinds* (what sets *individual* kinds apart is that in each possible world at most one thing instances them): "... since St. Thomas could maintain that every separate intelligence differed in kind from every other, what evil will there be in saying the same of every person and in conceiving individuals as final species" [G. R. Montgomery, ed., *Discourse on Metaphysics, Correspondence with Arnauld, Monadology* (La Salle, Ill.: Open Court, 1908), p. 237]. Now just as the essences of general kinds can be comparable, so can the essences of an individual and a general kind (*Bossie's* essence includes that of *cow*). But then why should the essences of individual kinds not be comparable too? There is every reason to see the relation between the Shroud of Turin and the piece of cloth as continuous with that between cow and animal: just as it is harder to be a cow than an animal, it is harder to be the Shroud of Turin than the piece of cloth, and just as nothing can be a cow without being an animal (but not conversely), nothing can be the Shroud of Turin without being the piece of cloth (but not conversely). So there seems to be a strong case for extending the familiar doctrine that the essence of one kind can include that of another to individual kinds, and, what comes to the same, to individuals themselves. (Incidentally, I am assuming that the Shroud of Turin could not have been made of anything other than that piece of cloth. Something made of another piece of cloth might have been called "the Shroud of Turin," but it would not have been our Shroud of Turin.)

certainly not in the *Shroud* of Turin's essence, the inclusion is lost. Equivalently, it ought to be possible to start with the essence of the Cloth of Turin, *add* the property of having served as the burial shroud of Jesus Christ (along perhaps with others this entails), and wind up with the essence of the Shroud of Turin. But, if the property of being identical to the Cloth of Turin is allowed into the Cloth of Turin's essence, then adding the property of having served as Jesus's burial shroud produces a sort of contradiction; for, obviously, *nothing* is both identical to the Cloth of Turin and necessarily possessed of a property—having served as Jesus's burial shroud—which the Cloth of Turin possesses only contingently. And the argument is perfectly general: if identity properties (or others like them) are allowed into things' essences, then distinct things' essences will always be incomparable.<sup>8</sup>

Implicit in the foregoing is a distinction between two types of property. On the one hand, there are properties that can only “build up” the essences in which they figure. Since to include such properties in an essence is not (except trivially) to keep any other property out, they will be called *cumulative*. On the other hand, there are properties that exercise an inhibiting effect on the essences to which they belong. To include this sort of property in an essence is always to block the entry of certain of its colleagues. Properties like these—identity properties, kind properties, and others—are *restrictive*. If restrictive properties are barred from essences, that will ensure that essences are comparable, and so preserve the intuition that each essence specifies what it takes to be the thing that has it.

Essences constrained to include only cumulative properties will have two advantages. First, they will determine their possessors' inessential properties negatively, not by what they include but by what they leave out; and, as a result, things' essences will be amenable to expansion into the larger essences of things it is “more difficult to be,” thus preserving the intuition that a thing's essence specifies what it takes to be that thing. And, second, things will be the things they are *in virtue of* having the essences they have. To put it approximately but vividly, they will be what they are because of what they are like.<sup>9</sup> Our tactic will be to look first for properties suited to inclusion in cumulative

<sup>8</sup> Identity properties are by no means the only properties that lead to these difficulties. Kind properties, for example, are just as bad. If the property of being a piece of cloth (i.e., being of the *kind* piece of cloth) is included in the Cloth of Turin's essence, then adding on the property of having served as Jesus's burial shroud (along with perhaps some others) can no longer yield the essence of the Shroud of Turin. For it is never essential to any piece of cloth that it should have been used in any particular way (necessarily, any piece of cloth could have been destroyed moments after its fabrication). Incidentally, kind properties are disqualified by the first argument too: like identity properties, they are possessed not *simpliciter*, but dependently on other properties. It defies credulity that two things should be indiscernible up to this detail, that one is a collie and the other is not. (Thinking of identity and kind properties as *classificatory*, rather than *characterizing*, the above becomes the truism that a thing's classification depends entirely on what the thing is like.)

<sup>9</sup> Although this is probably overstating it, at least as far as what is actually established goes (see Prop. 4 below). See also David Wiggins, *Sameness and Substance* (Cambridge, Mass.: Harvard, 1980), and Robert Merrihew Adams, “Primitive Thisness and Primitive Identity,” *Journal of*



essences and then to show that, under reasonable further assumptions, identity supervenes on cumulative essence.

### III. MODELING ESSENCE

To find a set of properties suitable for the construction of cumulative essences, one needs to know what “properties” are; especially because a totally unrestricted notion of property is incoherent, as Richard’s and Grelling’s paradoxes show.<sup>10</sup> So it makes sense to look for a sharper formulation of the notion of property before pushing ahead with the search for cumulative essence. Such a formulation is provided by the apparatus of possible worlds.

Let  $\mathcal{L}$  be an ordinary first-order language with identity, and let  $\mathcal{L}(\Box)$  be  $\mathcal{L}$  supplemented with the sentential necessity operator ‘ $\Box$ ’. To a first approximation, a model of  $\mathcal{L}(\Box)$  is just a set  $\mathcal{W}$  of models  $W$  of  $\mathcal{L}$  (to be thought of as possible worlds). But there is a qualification. Traditionally, a model’s domain is simultaneously the set of things that *can be talked about* and the set of things that *exist*, i.e., the domain of *discourse* and the *ontological* domain. But, since one can talk about things that do not exist,  $W$ ’s domain of discourse should be allowed to contain things not in its ontological domain; and since there are not, mystical considerations to the side, things about which one cannot talk,  $W$ ’s ontological domain should be a subset of its domain of discourse. What this means formally is that with each model  $W$  in  $\mathcal{W}$  is associated a subset  $\mathcal{D}(W)$  of its domain (intuitively, the set of things existing in  $W$ ). Let  $W$  thus supplemented be known as a *free model* of  $\mathcal{L}$ . For simplicity’s sake, every member of  $\mathcal{W}$  will have the same domain  $\mathcal{D}$ , and  $\mathcal{D}$  will be the union of the  $\mathcal{D}(W)$ ’s. And now a model of  $\mathcal{L}(\Box)$  can be defined as a set  $\mathcal{W}$  of free models of  $\mathcal{L}$ , such that the domain of discourse of each is the union of all their ontological domains.<sup>11</sup>

Tempting though it is to define a property as any function  $P$  from worlds  $W$  to subsets  $P(W)$  of  $\mathcal{D}$ , there is reason not to. For when will  $a$  have  $P$  necessarily:

*Philosophy*, LXXVI, 1 (January 1979): 5–26, in both of which the sufficiency of “quality” for “quiddity” is considered and rejected. Wiggins’s opinion is that

... to make clear which thing a thing is, it is not enough (*pace* the friends of the logically particularized essence) to say however lengthily that it is *such*, or *so and so*. We have to say that it is *this* or *that such*. This is perfectly obvious when we think of trying to determine one entity by mentioning short or simple predicates (other than *identical with x* or such like). But it is difficult to see any reason to believe that by making ordinary predicates ever longer and more complicated we shall be able to overcome the obvious non-sufficiency or non-necessity for identity with just  $x$  that infects all the relatively simple predicates true of  $x$  (104/5).

<sup>10</sup> See Peter Geach, “Identity,” in his *Logic Matters* (Oxford: Blackwell, 1972). Baruch Brody asserts that his theory, according to which items are identical if and only if they are indiscernible over “all” properties, “is not ruled out by its leading to any paradoxes” (*Identity and Essence* (Princeton, N.J.: University Press, 1980), p. 18). But he does not satisfactorily explain why not.

<sup>11</sup> The accessibility relation is omitted; in effect, every world has access to every other.



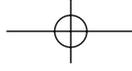
when it has  $P$  in every world, or when it has it in every world in which it exists? Not the former, because then everything necessarily exists.<sup>12</sup> Nor the latter, first, because it permits a thing to possess only accidentally a property it must perish to lose and, second, because it upsets the principle that essence varies inversely with existence, i.e., the fewer the worlds a thing exists in, the more properties it has essentially. What this in fact points up is a difference between two kinds of characteristic: being human in every world where you exist is sufficient for being human everywhere (almost all characteristics are like this), but existing in every world where you exist is obviously not sufficient for existing everywhere (apparently only existence and characteristics involving existence are like this). From now on, an *attribute* is a function from worlds  $W$  to subsets of  $\mathcal{D}$ , and a *property* is an attribute  $P$  such that anything having it wherever it exists has it everywhere. In general, an attribute is necessary to a thing if it attaches to the thing in every possible world (preserving the intuition that existence is sometimes contingent). If the attribute is also a property, this reduces to the thing's having the attribute wherever it exists (preserving the intuition that humanity is necessary to Socrates if he cannot exist without it). In what follows, properties (rather than attributes in general) are the items under investigation.

From the definition of property, it follows that, if  $P$  is a property, then so are  $P^\square: W \rightarrow \{a \in \mathcal{D} | \forall W' P(W')\}$  (the property of being essentially  $P$ , or  $P$ 's *essentialization*);  $P^\diamond: W \rightarrow \{a \in \mathcal{D} | \exists W' a \in P(W')\}$  (the property of being possibly  $P$ , or  $P$ 's *possibilization*); and  $P^\Delta: W \rightarrow \{a \in P(W) | \exists W' a \notin P(W')\}$  (the property of being accidentally  $P$ , henceforth  $P$ 's *accidentalization*). The essentialization  $X^\square$  (accidentalization  $X^\Delta$ ) of a set  $X$  of properties is the set of its members' essentializations (accidentalizations). If  $a$  is in  $P(W)$  and exists in  $W$ , then it is in  $P[W]$  (note the square brackets). If for each  $P$  in  $X$   $a \in P(W)$ , then  $a \in X(W)$ ; if, in addition,  $a$  exists in  $W$ , then it is in  $X[W]$ . A set  $Y$  of properties is *satisfiable* in  $W$ , written  $\text{Sat}[Y, W]$ , iff there is something in  $\bigcap_{P \in Y} P[W]$ . Given a set  $X$  of properties, a thing  $a$ 's *X-essence*  $\mathbb{E}_X(a)$  is the set of all  $P$  in  $X$  which  $a$  possesses essentially, or  $\{P \in X | \exists W (a \in P^\square(W))\}$ .  $\beta$  is an *X-refinement* of  $a$ , written  $a \leq_X \beta$  —or just  $a \leq \beta$  if  $X$  is clear from context—iff  $a$ 's  $X$ -essence is a subset of  $\beta$ 's, i.e., if  $\mathbb{E}_X(a) \subseteq \mathbb{E}_X(\beta)$ .

That essences drawn from  $X$  should be amenable to expansion is a condition not on  $X$  alone, but on  $X$  and  $\mathcal{W}$  taken together:  $X$  and  $\mathcal{W}$  must be so related that suitably expanding the  $X$ -essence of anything in any world in  $\mathcal{W}$  always produces the  $X$ -essence of some other thing existing in that same world. Let

<sup>12</sup> The problem existence raises for the definition of 'essential' is not unfamiliar. Kripke alludes to it in "Identity and Necessity":

Here is a lectern. A question which has often been raised in philosophy is: What are its essential properties? What properties . . . are such that this object has to have them if it exists at all . . . [Footnote:] This definition is the usual formulation of the notion of essential property, but an exception must be made for existence itself: on the definition given, existence would be trivially essential. We should regard existence as essential to an object only if the object necessarily exists. (86)



$\Omega = \langle \mathcal{W}, X \rangle$  be a *property-model* of  $\mathcal{L}(\square)$  if  $\mathcal{W}$  is a model of  $\mathcal{L}(\square)$  and  $X$  is a set of properties on  $\mathcal{W}$ . A property-model  $\Omega$  is *upward-closed*, or *u-closed*, iff:

$$(U) \quad \forall a \forall Y \subseteq X - \mathbb{E}_x(a) \forall W [a \in Y^\Delta[W] \Rightarrow (\exists \beta \geq a) \beta \in Y^\square[W]]$$

≡

In words, given any  $a$ , given any set  $Y$  of properties not essential to  $a$ , and given any world  $W$ , if  $a$  exists in  $W$  and has  $Y$  there, then it has a refinement  $\beta$  which exists in  $W$  and has  $Y$  essentially there. (For future reference, (U) is provably equivalent to the simpler statement that  $\forall Z \subseteq X \forall W [\text{Sat}[Z, W] \Rightarrow \text{Sat}[Z^\square, W]]$ .)

Upward closure requires that any existing  $a$  possessing (suitable) properties inessentially be refinable into an existing  $\beta$  that possesses those properties essentially. The converse is intuitive too: if there exists a  $\beta$  refining  $a$  which essentially possesses (suitable) properties not essential to  $a$ , then  $a$  should exist and possess those same properties accidentally. Thus, if the Shroud of Turin exists in a world  $W$ , then not only should the Cloth of Turin exist in  $W$ , but it should serve as Jesus's burial shroud in  $W$ . Not only is this plausible on the face of it, but otherwise it is hard to see what separates the worlds in which the Cloth occurs by itself from those in which it occurs together with the Shroud; whereas surely the difference is that in the latter, but not the former, the Cloth serves as Jesus's burial shroud. And, in general, if  $\beta$  refines  $a$ , then surely what separates worlds in which  $a$  exists without  $\beta$  from those in which  $a$  exists with  $\beta$  is that, in the latter worlds,  $a$  possesses the difference between their  $X$ -essences, whereas in the former it does not. Specifically, if  $\beta$  refines  $a$ , then (1)  $a$  exists wherever  $\beta$  does, (2) in worlds where both exist,  $a$  accidentally possesses every property in  $\mathbb{E}_x(\beta) - \mathbb{E}_x(a)$ , i.e., every property in the difference between their essences, and (3) in worlds where just  $\beta$  exists,  $a$  does not possess all the properties in  $\mathbb{E}_x(\beta) - \mathbb{E}_x(a)$ . All of these follow on the addition of a requirement of *downward closure*, literally the converse of the upward closure enforced above:

$$(D) \quad \forall a \forall Y \subseteq X - \mathbb{E}_x(a) \forall W [(\exists \beta \geq a) \beta \in Y^\square[W] \Rightarrow a \in Y^\Delta[W]]$$

≡

In words, for anything  $a$ , any properties  $Y$  not essential to it, and any world, if  $a$  has an existing refinement  $\beta$  possessing  $Y$  essentially, then  $a$  exists and possesses  $Y$  accidentally. Property-models satisfying both (U) and (D) are *closed*.

*Prop. 1* Let  $\Omega$  be closed. If  $a \leq \beta$ , then

$$(1) \quad \mathcal{W}(\beta) \subseteq \mathcal{W}(a)$$

$$(2) \quad \forall W \in \mathcal{W}(\beta) [a \in (\mathbb{E}_x(\beta) - \mathbb{E}_x(a))^\Delta(W)]$$

$$(3) \quad \forall W \in \mathcal{W}(a) - \mathcal{W}(\beta) [a \notin (\mathbb{E}_x(\beta) - \mathbb{E}_x(a))^\Delta(W)]$$

≡

≡

*Proof:* For (1), observe first that  $\Lambda[W] = \Lambda(W) \cap \mathcal{D}(W) = \mathcal{D} \cap \mathcal{D}(W) = \mathcal{D}(W)$  (because the null intersection is everything, in this case  $\mathcal{D}$ ). By (D),  $\beta \in \mathcal{D}(W) \Rightarrow \beta \in \Lambda[W] \Rightarrow \beta \in \Lambda^\square[W] \Rightarrow a \in \Lambda^\Delta[W] \Rightarrow a \in \Lambda[W] \Rightarrow a \in \mathcal{D}(W)$ .  $\square$





For (2), just let  $Z$  be  $\mathbb{E}_x(\beta) - \mathbb{E}_x(\alpha)$ . For (3), suppose that  $\beta$  does not exist in  $W$ , and suppose *per absurdum* that  $\alpha$  accidentally possesses, in  $W$ , the difference between its  $X$ -essence and  $\beta$ 's. By upward closure,  $\alpha$  has a refinement  $\gamma$  which exists in  $W$  and which possesses the whole lot, i.e., all of  $\mathbb{E}_x(\beta)$ , essentially. But then  $\gamma$  refines  $\beta$ ; so, by (1),  $\beta$  exists in  $W$  after all. Contradiction.  $\square$

≡

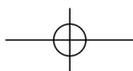
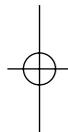
≡

#### IV. CONTINGENT IDENTITY

Things that disagree in any of their properties are not identical. The Shroud of Turin, which (let us suppose) *had* to enshroud Jesus, is thus distinct from the Cloth of Turin, which did not. But, as we said, there is something deeply troubling about leaving matters thus. After all, the Shroud of Turin is also distinct from the Treaty of Versailles. Do we really want to leave the Shroud's relations with the Cloth on the same level as its relations with the Treaty of Versailles? And the trouble does not stop here. The Cloth and the Shroud differ, it is true, but it must also be said that their differences are of a somewhat *recherché* variety. In every *ordinary* respect the two are exactly alike. And this is on the face of it a rather extraordinary coincidence. That the Cloth and the Shroud are specially connected seems undeniable, but something must be done to demystify the connection. If it is not identity, what is it?

Maybe the answer is that it *is* identity, but identity of a different, less demanding, character. In the terms of a currently unpopular theory—and notwithstanding the argument that seems to rule it out—it is “contingent identity,” or (the more neutral term) “coincidence.”<sup>13</sup> Despite the once widespread enthusiasm for contingent identity, it seems to me that the idea never received a satisfactory formulation. Specifically, all the analyses I have seen have a drawback in common: they (sometimes explicitly, sometimes in effect) treat things as strung together out of their modal manifestations (states, slices, stages), and call them coincident in a world if their manifestations in that world are properly identical. There are two objections to this kind of explication. The first is that it relies, ultimately, on the notion to be explicated; for one has little idea what a thing's state or manifestation in a world is, if not something whose nature is exhausted by its being exactly like the thing so far as the relevant world is concerned, i.e., by its being contingently identical with the thing in that world. Even more important, intuitively, things (e.g., animals) are *not* strung together out of their modal manifestations in this way, and proper identity of modal manifestations is

<sup>13</sup> Another reason for preferring “coincident” to “contingently identical” is that properly identical things will also be coincident, indeed necessarily so, and it sounds funny to say that they are *necessarily* contingently identical. But I continue to use the term, partly for shock value, and partly for reasons to be given presently.



not what is meant by contingent identity. Intuitively, things are just, well, *things*, and coincidence is a matter of things' circumstantial sameness.

How, then, is circumstantial sameness to be separated out from total sameness? What marks off the "ordinary" respects in which the Cloth and the Shroud are alike from the "extraordinary" respects in which they differ? Let us start with Dana Scott's idea that "two individuals that are generally distinct might share all the same properties (of a certain kind!) with respect to the present world." ("Advice on Modal Logic," *op. cit.*, p. 165). Probably the most obvious way of elucidating this would be to say the following:  $\alpha$  and  $\beta$  are contingently identical (in a world) if and only if they have the same *contingent properties* (in that world). So, for example, the bust and the hunk of wax agree in their size, weight, color, and so on—and all these are, of course, properties they have contingently. Maybe contingent identity is sameness of contingent properties.

That that cannot be right follows from the fact that, if anything has a property contingently, then it has all its stronger properties contingently too. (Suppose  $\alpha$  has  $P$  contingently; then there is a world in which  $\alpha$  lacks  $P$ ; but if  $Q$  is stronger than  $P$ ,  $\alpha$  also lacks  $Q$  there; and, since  $\alpha$  has  $Q$  it has it contingently.) So, for example, if Paris is only contingently romantic, then it is only contingently identical-to-Paris-and-romantic. But that means that anything that has the same contingent properties that Paris has is (among other things) identical with Paris. And that already shows that Paris is the only thing with exactly its contingent properties. Thus, if contingent identity is treated as sameness of contingent properties, contingent identity collapses into identity proper.

Still, from a certain perspective, this first analysis might be only a little way off the mark. To the question, What makes a thing's possession of a property *circumstantial*? it seems natural to reply that the possession is circumstantial if it depends on *how matters actually stand with the thing*. But now notice that this is ambiguous. Depends *how: partly* or *wholly*? If you answer "partly," then you get the thing's *contingent* properties. But, if you answer "wholly," you get the properties the thing has *entirely* in virtue of how matters actually stand with it; and these properties, what can be called the *categorical* properties, seem intuitively to be the ones in question.<sup>14</sup> For if two things agree in all their categorical properties (in a world), then *so far as that world and it alone* is concerned, the two things are just the same. And that is what was meant by "contingent identity." So contingent identity is categorical indiscernibility.

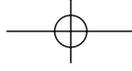
To come to this conclusion from a different direction, consider again the driving home and the speeding home. What separates the "ordinary" respects in which these two are alike from the "extraordinary" respects in which they differ? For a start, the driving home could have been done slowly, but not so the

<sup>14</sup> To be absolutely clear about the difference between contingency and categoricity, consider their complements. Where a thing has its noncontingent properties *necessarily*, it has its noncategorical properties *hypothetically*.

speeding home; the driving home had higher prior probability than the speeding home; if the driving home had not occurred, Jones would have taken the bus home, but the same cannot be said of the speeding home; and the speeding home, rather than the driving home as such, caused Jones's accident. Thus, the driving home and the speeding home differ in—among other things—their modal, probabilistic, counterfactual, and casual properties. Now what is special about modal, probabilistic, counterfactual, and casual properties? Primarily this: they are grounded not just in how a thing *actually* is, but on how it *would* or *could* have been if circumstances had been different. All a thing's other properties, by contrast, are grounded entirely in how it is in the circumstances that happen to obtain. The former properties are a thing's *hypothetical* properties, the latter its *categorical* properties. Now the contingent identity of the driving home with the speeding home seems intuitively to be a matter of their sharing such properties as speed, place, time, etc., regardless of their modal, causal, probabilistic, and counterfactual differences; that is, their contingent identity seems to be a matter of their sharing their categorical properties, irrespective of their hypothetical differences.

How are a thing's categorical properties to be found? (Actually it will be simplest if we look for properties categorical *as such*, i.e., properties that can only be had in a manner independent of what would or could have happened.) Why not take the intuitive notion that a property is categorical just in case a thing's having it is independent of what goes on in nonactual worlds, and try to turn this into a definition? The problem is that such a definition would be circular. Suppose it is a categorical property of this hunk of clay that it is spherical. How can that depend on how the clay comports itself in other worlds? But, if you think about it, it does, in that the clay's being spherical in this world depends on its being, in those worlds, such that in *this* world it is spherical. The problem is that being such that, in this world, it is spherical, is a hypothetical property of the clay. So, apparently, what we really meant to say was that a property is categorical if it attaches to a thing regardless of its *categorical properties* in other worlds. And that is clearly circular.<sup>15</sup>

<sup>15</sup> Two remarks. As a characterization of the categorical properties, the foregoing is circular. But it does have the virtue of illustrating why not *every* property can be hypothetical (as is sometimes suggested). For a property to be hypothetical, whether a thing has it must depend on the things' *categorical* properties in other worlds; and that shows that no property can be hypothetical unless at least some properties are categorical. And, since it is relatively unproblematic that categorical properties give rise to hypothetical properties (given the present broadly essentialist assumptions), neither category can be emptied without emptying the other. So a skeptic about the distinction should maintain that no property is of *either* kind, not that all (some) properties are of one kind and none are of the other. (For example, Sydney Shoemaker's theory of properties as "second-order powers," though it might seem to imply that all properties are hypothetical, or, on another reading, that all properties are categorical, is perhaps better read as rejecting the distinction altogether. See "Causality and Properties," and "Identity, Properties, and Causality," collected in his *Identity, Cause, and Mind* (New York: Cambridge, 1984).) Second, in rejecting the proposed account of categoricity on grounds of circularity, I do not mean to imply that the account I finally give is not itself ultimately circular. Given the cumulative properties, the categorical properties can be



Somehow the circularity has to be circumvented. Things are going to be coincident in a world iff they have exactly the same categorical properties there. But maybe this can be turned around: the categorical properties are exactly those which cannot tell coincident things apart.

Postpone for a moment the question of how that would help, and ask, instead, is it even true? That is, *are* the categorical properties the properties insensitive to the difference between coincidents? This will be the case only if coincidence is compatible with every kind of hypothetical variation, i.e., if, for every hypothetical property, coincidents can be found that disagree on it. But it is clear that ordinary things do not exhibit the hypothetical variety that this would require. Among ordinary things, coincidents never differ on (e.g.) the score of fragility (if the statue is fragile then so is the piece of clay); among ordinary things, one never finds one thing accidentally juvenile, or mature, and another, coincident with the first, essentially so (simply because *no* ordinary thing is essentially juvenile, or mature).

So much for ordinary things. But what about things as seen from the vantage point of metaphysics? Metaphysics aspires to understand reality as it is in itself, independently of the conceptual apparatus observers bring to bear on it. Even if we do not ourselves recognize essentially juvenile or mature entities, it is not hard to imagine others who would;<sup>16</sup> and to someone who, in addition to the statue and the piece of clay, discerned a statue-cum-shards, not everything coincident with the statue would be fragile. Conversely, we recognize things, say, essentially suitable for playing cribbage, or cutting grass, which others do not, or might not have. To insist on the credentials of the things *we* recognize against those which others do, or might, seems indefensibly parochial. In metaphysics, unusual hypothetical coloring can be no ground for exclusion.<sup>17</sup> Since this is metaphysics, everything up for recognition must actually be recognized; and, when this is done, there are coincidents enough to witness the hypotheticality of every hypothetical property.<sup>18</sup>

noncircularly specified; but the cumulative properties themselves cannot be noncircularly specified, in particular not by the formal conditions laid down above.

<sup>16</sup> To get a sense of what it might be like to countenance a creature coming into existence “in mid-life,” consider Jane Eyre’s reflections on the eve of her (anticipated) marriage to Mr. Rochester: “Mrs. Rochester! She did not exist: she would not be born till tomorrow, sometime after eight o’clock A.M.; and I would wait to be assured she had come into the world alive before I assigned her all that property.” For a creature that stops existing “in mid-life,” there are the opening lines of Neil Young’s “A Child’s Claim to Fame”: “I am a child/I last a while.”

<sup>17</sup> Less dogmatically, there are two kinds of metaphysics: descriptive and transcendental. In descriptive metaphysics one is interested in reality as people see it; in transcendental metaphysics one tries to abstract to the largest extent possible from the human contribution. Pretty clearly, the distinction is relative. All metaphysics is somewhat transcendental (metaphysicians do not spend much time thrashing out the nature of time zones), but probably the present approach is more transcendental than most.

<sup>18</sup> To say that everything up for recognition is actually recognized, is to say the following: given any set of worlds, and given an assignment to each of categorical properties satisfiable therein, there





Given information about what was coincident with what, the categorical properties could be identified: they would be the properties insensitive to the difference between coincidents. Now, as of yet, there *is* no information about what is coincident with what (that is why we were looking for the categorical properties in the first place). But that is not to say that none can be obtained; and, in fact, *certain* cases of coincidence—enough to weed out all the noncategoricals—are discoverable in advance.

To find these cases, try to imagine pairs of things that differ as little as possible from being strictly identical (for things almost identical will be contingently identical if any things are). Trivially, if  $\alpha$  and  $\beta$  are strictly identical,  $\alpha$  will exist in exactly the same possible worlds as  $\beta$ , and  $\alpha$  will be coincident with  $\beta$  in all of them. To arrange for the least possible departure from this, let  $\alpha$  exist in a few more worlds than  $\beta$ , but otherwise leave everything unchanged, i.e., let them be coincident in all the worlds where both exist. As it happens, that is exactly how it is with the driving home and the speeding home. Wherever the speeding home occurs, the driving home occurs too, and is coincident with the speeding home. But there will also be worlds in which the driving home is done at a reasonable speed, and in such worlds the speeding home does not occur.

Still, none of this helps with the project of explicating contingent identity, unless there is a way of characterizing the given relation—the relation between the driving home and the speeding home—which does not itself rely on the notion of contingent identity. But there is: it is the relation of refinement. Although only a fraction of all coincidents stand in the relation of refinement, this fraction is enough to weed out all the noncategorical properties.<sup>19</sup> With the noncategorical properties weeded out, the categorical properties are isolated. And

is something that exists in those worlds exactly and possesses in each the associated categorical properties. Call this the requirement of *fullness*.

<sup>19</sup> To see *why* the properties insensitive to the difference between things related by refinement can be relied on to be exactly the categorical: Call these properties the *provisionally* categorical properties, and call things indiscernible with respect to these properties *provisionally* coincident. The problem is really to prove that every provisionally categorical property is genuinely categorical (the converse is clear). Let  $P$  be provisionally categorical. Then, by the definition of provisional coincidence,  $P$  cannot distinguish provisionally coincident things. Suppose toward a contradiction that  $P$  is not genuinely categorical. Then there are  $\alpha$  and  $W$  such that  $\alpha$  possesses  $P$  in  $W$ , but its possession of  $P$  in  $W$  depends on what worlds (other than  $W$ ) it inhabits or on its genuinely categorical properties in those worlds. Thus, if there were something genuinely coincident with  $\alpha$  in  $W$ , but differing from  $\alpha$  in the worlds (other than  $W$ ) it inhabited, or the genuinely categorical properties it had in them, that thing would lack  $P$  in  $W$ . Specifically, something existing in worlds  $W, W', W'' \dots$ , and possessing the genuinely categorical properties  $Y$  in  $W$  ( $Y$  is the set of  $\alpha$ 's genuinely categorical properties in  $W$ ),  $Y'$  in  $W'$ ,  $Y''$  in  $W'' \dots$ , would lack  $P$  in  $W$ . To produce such a thing, let  $\gamma (= \alpha), \gamma', \gamma'' \dots$  be entities satisfying  $Y, Y', Y'', \dots$  in  $W, W', W'' \dots$ . If  $Z, Z', Z'' \dots$  are the sets of provisionally categorical properties possessed by  $\gamma, \gamma', \gamma'' \dots$  in  $W, W', W'' \dots$ , then, by fullness (see the preceding note), there is a  $\beta$  existing in exactly  $W, W', W'' \dots$ , and possessing  $Z$  in  $W, Z'$  in  $W', Z''$  in  $W'' \dots$ . Since a thing's provisionally categorical properties include its genuinely categorical properties,  $\beta$  meets the conditions laid down above for lacking  $P$  in  $W$ . But, by its definition,  $\beta$  is provisionally coincident with  $\alpha$  in  $W$ . Since  $P$  distinguishes provisional coincidents, it is not provisionally categorical after all.



with the categorical properties in hand, contingent identity is at last explicable: things are contingently identical in a world if they have the same categorical properties there.

Trivial cases aside, contingently identical items will not be identical as a matter of necessity. But then what about the argument that purported to show that identities obtained necessarily if at all? Was the argument invalid? No; it showed that *something* was impossible. The question is, was the refuted possibility really that of contingent identity? Looking back at the argument, the crucial assumption was this: to be contingently identical, things have to have *all* their properties, up to and including properties of the form *necessary identity with such-and-such*, in common. If that is right, then contingent identity is, as argued, impossible. So the question is, *do* contingently identical things have to have all their properties, not only categorical but hypothetical as well, in common?

They do not. To agree that they did would be to concede the very point of contingent identity and to frustrate the clear intent of its advocates, which was that it was to be a relation *compatible* with counterfactual divergence. Understanding contingent identity as sameness of nonhypothetical properties, on the other hand, preserves its point and sustains it against the “proof” of its impossibility. Still, why did it even *seem* that contingent identity entailed absolute indiscernibility? Probably because it was taken for granted that contingently identical things were (at least) *properly* identical, only—and this was their distinction—*not necessarily so*.<sup>20</sup> (And the expression ‘contingent identity’ can certainly be faulted for encouraging this interpretation.) Admittedly, contingent identity in *this* sense is not possible.<sup>21</sup> But there is a better and more generous way of understanding contingent identity: strict and contingent identity are different relations, and, because of their differences as relations, one can obtain contingently whereas the other cannot. It only remains to spell out the formal details.

<sup>20</sup> True, Smart does go out of his way to emphasize that “the brain-process doctrine asserts identity in the *strict* sense” (“Sensations and Brain Processes,” *op. cit.*, p. 145). But by this he seems to mean that he is not talking about the relation that one thing bears to another when they are “time slice[s] of the same four-dimensional object” or when they are “spatially or temporally continuous” (145). Certainly there is nothing to suggest that he had in mind a contrast between “strict identity” and coincidence. What is clear is that he took “strict identity” to be a relation fully compatible with hypothetical dissimilarity.

<sup>21</sup> Although some philosophers would say that identity *itself* can obtain contingently. These philosophers can *to some extent*, be seen as questioning the notion that is here called “identity proper” and as taking something *roughly* analogous to what is here called “coincidence” to be all the identity there is. Since this relation, which, relative to their schemes, probably deserves to be called “identity proper,” can obtain contingently, the kind(s) of contingent identity they advocate is (are) in a certain sense more radical than the kind assayed here. See, for example, Allan Gibbard, “Contingent Identity,” *Journal of Philosophical Logic*, IV, 2 (May 1975): 187–221; David Lewis, “Counterpart Theory and Quantified Modal Logic,” *Journal of Philosophy*, LXV 5 (March 7, 1968): 113–126, and *Counterfactuals* (Cambridge, Mass.: Harvard, 1973); and Robert Stalnaker, “Counterparts and Identity,” *Midwest Studies in Philosophy*, XI (Minneapolis: Minnesota UP, 1986).

## V. MODELING CONTINGENT IDENTITY

Formally, a property  $P$  is *categorical* iff necessarily, if  $\alpha$  and  $\beta$  are related by refinement, then  $\alpha$  has  $P$  iff  $\beta$  does, i.e., iff

$$(\forall W)(\forall \alpha, \beta \in \mathcal{D}(W))(a \leq \beta \Rightarrow (\alpha \in P(W) \Leftrightarrow \beta \in P(W))).$$

What is the relation between the cumulative properties and the categorical properties? Closure implies a partial answer.

*Prop. 2* If  $\Omega$  is closed, then every cumulative property is categorical.

*Proof:* Let  $P \in X$ , and let  $\alpha, \beta \in \mathcal{D}(W)$ ,  $\alpha \leq \beta$ . By u-closure, there are  $\alpha^*$  and  $\beta^*$  in  $\mathcal{D}(W)$  such that  $\forall P \in X (\alpha \in P[W] \Rightarrow \alpha^* \in P^\square[W])$  and  $\forall P \in X (\beta \in P[W] \Rightarrow \beta^* \in P^\square[W])$ . We show that  $\alpha \in P[W] \Leftrightarrow \beta \in P[W]$ . Since  $\alpha \leq \beta \leq \beta^*$ ,  $\alpha \leq \beta^*$ . Therefore,  $\beta \in P[W] \Rightarrow \beta^* \in P^\square[W] \Leftrightarrow \alpha \in P[W]$  (by d-closure). For the converse, notice first that  $\beta \leq \alpha^*$ . For if  $Q \in \mathbb{E}_X(\beta)$ , then, by d-closure,  $\alpha \in Q[W]$ ; whence  $Q \in \mathbb{E}_X(\alpha^*)$ . Since  $\beta \leq \alpha^*$ ,  $\alpha \in P[W] \Rightarrow \alpha^* \in P^\square[W] \Rightarrow \beta \in P[W]$  (by d-closure).  $\square$

Things are *coincident* in  $W$ —written  $\alpha \approx_W \beta$ —if they have the same categorical properties there. But, for the definition of categoricity to achieve its purpose, the system of coincidents has to be *full* or *complete*. Informally, this means that every point in the logical space of possible coincidents must actually be occupied; formally, for any (partial) function  $f$  from worlds  $W$  to things existing in  $W$ , there is a thing existing, and coincident with  $f(W)$ , in exactly the worlds in  $f$ 's domain.  $\Omega$  is *full* if it satisfies

$$(F) \forall f: W \in \mathcal{W} \rightarrow f(W) \in \mathcal{D}(W) \quad \exists \alpha [\mathcal{W}(\alpha) = \text{dom}(f) \ \& \ \forall W \in \mathcal{W}(\alpha) \ \alpha \approx_W f(W)]$$

Prop. 2 showed that, if  $\Omega$  is closed, then every cumulative property is categorical. For the converse, let  $\Omega$  be *maximal closed* if

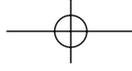
$$(M) \Omega = \langle \mathcal{W}, X \rangle \text{ is closed, and there is no } X' \text{ extending } X \text{ such that } \langle \mathcal{W}, X' \rangle \text{ is closed.}$$

If  $\Omega$  is maximal closed and full, then every categorical property is cumulative. In other words, a property is cumulative if and only if it is categorical.

*Prop. 3* Let  $\Omega$  be maximal closed and full. Then

$$\forall P [P \text{ is cumulative} \Leftrightarrow P \text{ is categorical}]$$

*Proof:*  $[\Rightarrow]$  This is just Prop. 2.  $[\Leftarrow]$  Let  $P$  be categorical. I claim that  $\langle \mathcal{W}, X+P \rangle$  is closed. For u-closure, let  $Z \subseteq X$  and suppose that  $\alpha$  has  $Z+P$  in  $W$ . By fullness, there is an  $\alpha^*$  which exists in  $W$  only and which is



coincident (with respect to  $X$ ) with  $a$  in  $W$ . Since  $a^*$  exists in  $W$  only, it has all its properties essentially. In particular,  $a^*$  has all of  $a$ 's categorical (w.r.t.  $X$ ) properties essentially, and so (by Prop. 2)  $a^*$  refines  $a$  (w.r.t.  $X$ ). Since  $P$  is categorical (w.r.t.  $X$ ),  $a^*$  has  $P$  in  $W$  too, and so it has  $P$  essentially in  $W$ . Thus  $a^*$  has  $(Z+P)^\square$  in  $W$ . For d-closure, let  $Z \subseteq X$  and let  $a \leq \beta \in (Z+P)^\square[W]$ . By the d-closure of  $\Omega$ ,  $a \in Z[W]$ , and, since  $P$  is categorical (w.r.t.  $X$ ),  $a \in P[W]$  too. Since  $\langle \mathscr{W}, X+P \rangle$  is both u-closed and d-closed,  $P$  is in  $X$ .<sup>22</sup>  $\square$

## VI. ESSENCE AND IDENTITY

Can distinct things have the same cumulative essence? So far, nothing prevents it. For example, there is nothing to rule out the following: there are exactly two possible worlds,  $W$  and  $W'$ ;  $a$  and  $\beta$  exist in both worlds;  $\gamma$  exists in  $W$  alone; and  $\gamma'$  exists in  $W'$  alone. In this situation  $\gamma$  will have to refine both  $a$  and  $\beta$  in  $W$ , and  $\gamma'$  will refine both  $a$  and  $\beta$  in  $W'$ . From the fact that  $a$  and  $\beta$  have a common refinement in each world, it quickly follows that they are coincident, i.e., have the same categorical properties, in each world. But, by Prop. 3, the categorical properties are exactly the cumulative properties; and, if  $a$  and  $\beta$  share their cumulative properties in every world, how can they have different cumulative essences?

Actually, this raises the critical question, avoided until now, of how coincidence and identity are related. Can  $a$  and  $\beta$  exist in the same worlds, be coincident in all of them, and still be distinct? It is hard to imagine how they could. For presumably, distinct items differ in one or another of two ways. Either they exist in different worlds, or they exist in the same worlds and are unlike, i.e., have different categorical properties, in at least one of them. Between distinct things, that is, there have got to be either intra-world or extra-world differences.

But what is the argument for this? If things exist in the same worlds, then, unless they have different categorical properties in at least one of them,

<sup>22</sup> Although it is wrong to try to *explicate* contingent identity in terms of identity of modal states (the notion of modal state depending on that of contingent identity), Prop. 3 suggests a way to define modal states so that the equation comes out true. Call  $a^*$   $a$ 's state in  $W$  iff  $a^*$ 's cumulative essence is exactly the set of  $a$ 's categorical properties in  $W$ . To see that if  $a$  exists in  $W$ , its state in  $W$ , exists too: By fullness, there is a  $\beta$  existing in  $W$  alone and coincident with  $a$  there. Thus  $\beta$ 's categorical properties in  $W$  are exactly  $a$ 's. By Prop. 3,  $\beta$ 's cumulative properties in  $W$  are exactly  $a$ 's categorical properties in  $W$ . Since  $\beta$  has all its properties essentially,  $\beta$ 's cumulative essence is the set of  $a$ 's categorical properties in  $W$ . Now it is easy to verify that things are coincident in a world iff they have strictly identical states there:  $a$  coincides with  $\beta$  in  $W$  iff  $a$  and  $\beta$  have the same categorical properties in  $W$  iff anything whose cumulative essence is the set of  $a$ 's categorical properties is also something whose cumulative essence is the set of  $\beta$ 's categorical properties iff any state of  $a$  in  $W$  is a state of  $\beta$  in  $W$ . (For the uniqueness of  $a$ 's state in  $W$ , see Prop. 4.)

the hypothesis of their distinctness can find no foothold. Take the standard example of “indiscernible” spheres afloat in otherwise empty space (suppose for argument’s sake that they exist in no other world). If the spheres were in *exactly the same place*, could they still be reckoned distinct? An hypothesis *so* gratuitous is beyond not only our powers of belief, but even our powers of stipulation. If, on the other hand, the spheres are in different places, then they differ on the (presumably categorical) properties of being in those places. (The properties have to be different, because they map the world in question to different spheres.)<sup>23</sup>

Call a property-model *separable* if it satisfies

$$(5) \quad \forall \alpha \forall \beta [(\mathcal{W}(\alpha) = \mathcal{W}(\beta) \ \& \ \forall W \epsilon \mathcal{W}(\alpha) \cup \mathcal{W}(\beta) \ \alpha \approx_w \beta) \Rightarrow \alpha = \beta]$$

The last proposition shows that, if  $\Omega$  is (besides being closed) separable, then things with the same cumulative essence are identical.

*Prop. 4* Let  $\Omega$  be closed and separable. Then

$$\forall \alpha \forall \beta [\mathbb{E}_X(\alpha) = \mathbb{E}_X(\beta) \Rightarrow \alpha = \beta]$$

*Proof:* Let  $\alpha$  and  $\beta$  have the same  $X$ -essence. Since  $\alpha$  and  $\beta$   $X$ -refine each other, by Prop. 1 they exist in the same worlds. By the definition of  $X$ -categoricity, in each of these worlds any  $X$ -categorical property attaching to either attaches also to the other. Thus  $\alpha$  and  $\beta$  have the same  $X$ -categorical properties, and so coincide, in every world where they exist. By separability,  $\alpha$  is identical with  $\beta$ .  $\square$

To this extent, essence determines identity.<sup>24</sup>

<sup>23</sup> Granted, one cannot identify these properties without appeal to the objects that have them; but the claim was that they have different categorical properties, not that one can distinguish them by their categorical properties. Granted also, except in connection with the world in question, the properties’ extensions will be to a large extent arbitrary (when is something in this world in the same place as the first sphere in that one?); but that does not matter, so long as the arbitrary choices are made in such a way that the resulting properties are categorical.

<sup>24</sup> Even if this result is accepted, there is room for doubt about its precise significance. Briefly, one worry is that to distinguish  $\alpha$ ’s cumulative essence from  $\beta$ ’s, one would already have to be able to distinguish  $\alpha$  from  $\beta$ . But this seems to confuse the metaphysical thesis that distinct things have different essences with the epistemological doctrine that distinct things can always be *distinguished* by their essences. Second, not everything here called a “property”—basically, functions from worlds to extensions—is a *genuine property*. If this means that genuine properties are not functions, this is granted; but it does not matter, if, for every world-to-extension function, there is some genuine property such that the function takes each world to the set of things possessing that property therein. But the criticism survives in the form: not all world-to-extension functions (not even all cumulative ones) *are* induced in this way by genuine properties. And that is undeniable. Further progress depends on figuring out what makes genuine properties genuine. (For more on the difference between genuine and pseudo properties, see the articles by Sydney Shoemaker mentioned in note 15 above.)

## VII. APPLICATIONS

(A) Treating contingent identity as sameness of categorical properties goes *part* of the way toward solving a problem David Wiggins raises for relative identity in *Sameness and Substance*. He argues there that, since (i) what sets identity relations apart from the common run of equivalence relations is their satisfaction of Leibniz's law, and (ii) no variety of relative identity can satisfy (an unrestricted version of) Leibniz's law, (iii) relations of relative identity are not identity relations (pending discovery of a suitably restricted form of Leibniz's law). To answer this argument, one would need an uncontrived law of the form: if  $a$  is the same  $f$  as  $\beta$ , then  $a$  and  $\beta$  have thus-and-such properties in common. However, Wiggins thinks such a law will prove impossible to formulate:

It seems that the very least we shall require is more information about the case of the *restricted* congruence that results from the  $g$ -identity, for *some one* sortal concept  $g$ , of  $x$  and  $y$ . No stable formulation of restricted congruence is available, however. Nor, I suspect, will it ever be given (39).

But, if contingent identicals are seen as the *same concrete thing*, then a rigorous restriction of Leibniz's law is at hand: if  $a$  and  $\beta$  are the same concrete thing, then they have the same categorical properties.<sup>25</sup>

(B) Nearly everyone's gut reaction to functionalism is that *phenomenal* properties, at any rate, cannot be functional, because nothing functional can attain to the "manifest" character of felt experience. Perhaps this idea finds support in the categorical/hypothetical distinction. The property of playing functional role  $R$  is the property of bearing certain complicated counterfactual relations to inputs, outputs, and the players of various other functional roles. Details aside, such a property is obviously hypothetical. But the property of painfulness (note: not the property of causing pain, but that of being pain) *seems* to be a categorical property *par excellence*. Therefore, painfulness is not a functional property. (Note that this affects only the version of functionalism that flatly identifies mental properties with functional properties.)

(C) That mental and physical events are not properly identical is argued not only by their essential differences (emphasized by Kripke), but also by their causal differences. Suppose Smith's pain is identical with neural event  $\nu$ , which causes neural event  $\epsilon$ . Then the strict identity theorist will have to say that Smith's pain caused  $\epsilon$ ; but that seems questionable, because  $\nu$ 's *being* Smith's pain contributed nothing to its production of  $\epsilon$  (one wants to say: even if  $\nu$

<sup>25</sup> Whether this helps with the general problem of saying what properties relative identicals must have in common is another question, but one perhaps worth exploring. See also Nicholas Griffin, *Relative Identity* (New York: Oxford, 1975), secs. 1.2 and 8.5.

had not been Smith's, or any, pain,  $\epsilon$  would still have eventuated). If Smith's pain and  $\nu$  are only coincident, on the other hand, then *naturally* they will have different causal powers and susceptibilities; and a sensitive counterfactual theory of causation might be able to *exploit* their essential differences to predict their causal differences.<sup>26</sup> An event  $a$  causes an event  $\beta$ , other things equal, only if it is *required* for  $\beta$  in the following sense: given any (actually occurring) event  $\gamma \approx a$  whose essence does not include  $a$ 's, if  $\gamma$  had occurred in  $a$ 's absence,  $\beta$  would not have occurred; and only if it is *enough* for  $\beta$  in the following sense: given any (actually occurring) event  $\gamma \approx a$  whose essence is not included in  $a$ 's,  $\gamma$  is not required for  $\beta$ . As for Smith's pain (call it  $\pi$ ), that  $\epsilon$  would have occurred even in  $\pi$ 's absence, provided that  $\nu$  had still occurred, shows that  $\pi$  is not required for  $\epsilon$ ; and that  $\nu$  is (let us assume) required for  $\epsilon$  shows that  $\pi$  is not enough for  $\epsilon$  either. Complementary considerations show how it can be Smith's decision, rather than the corresponding neural event, that causes her action.

## REFERENCES

- Adams, Robert Merrihew (1979). "Primitive Thisness and Primitive Identity". *Journal of Philosophy*, LXXVI, 1 (Jan.), 5–26.
- Brody, Baruch (1980). *Identity and Essence*. Princeton, N.J.: Princeton University Press.
- Geach, Peter (1972). "Identity". In Geach, *Logic Matters*. Oxford: Blackwell.
- Gibbard, Allan (1975). "Contingent Identity". *Journal of Philosophical Logic*, IV, 2 (May), 187–221.
- Griffin, Nicholas (1975). *Relative Identity*. New York: Oxford University Press.
- Kripke, Saul (1977). "Identity and Necessity". In Stephen P. Schwartz (ed.), *Naming, Necessity, and Natural Kinds*. Ithaca, N.Y.: Cornell.
- (1980), *Naming and Necessity*. Cambridge, Mass.: Harvard.
- Lambert, Karel (ed.) (1970), *Philosophical Problems in Logic*. Boston: D. Reidel.
- Lewis, David (1968). "Counterpart Theory and Quantified Modal Logic". *Journal of Philosophy*, LXV, 5 (7 Mar.), 113–26.
- (1973). *Counterfactuals*. Cambridge, Mass.: Harvard.
- Mason, H. T. (ed.) (1967) *The Leibniz–Arnauld Correspondence*. Manchester: Manchester University Press.
- Montgomery, G. R. (ed.) (1908), *Discourse on Metaphysics, Correspondence with Arnauld, Monadology*. La Salle, Ill.: Open Court.
- Shoemaker, Sydney (1984). *Identity, Cause, and Mind*. New York: Cambridge.
- Smart, J.J.C. (1959), "Sensations and Brain Processes". *Philosophical Review*, LXVIII, 2 (Apr.), 141–56.
- Stalnaker, Robert (1986). "Counterparts and Identity". *Midwest Studies in Philosophy*, XI. Minneapolis: Minnesota UP.
- Wiggins, David (1980), *Sameness and Substance*. Cambridge, Mass.: Harvard.

<sup>26</sup> Thanks to Barry Loewer for talking to me about this; he and Paul Boghossian are working along similar lines.



## 2

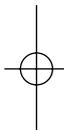
# Intrinsicness

### I. INTRODUCTION

*You* know what an intrinsic property is: it's a property that a thing has (or lacks) regardless of what may be going on outside of itself. To be intrinsic is to possess the second-order feature of stability-under-variation-in-the-outside-world.

You probably know too why this is hopeless as a philosophical account of intrinsicness. "Variation in the outside world" has got to mean "variation in what the outside world is like *intrinsically*." Otherwise every property G is extrinsic; for a thing cannot be G unless the objects outside of it are *accompanied* by a G.<sup>1</sup>

But, although the naive account is circular, it is not beyond salvage. Leave aside for a minute the problem of saying what *in general* constitutes intrinsic variation in the outside world. A special case of the notion can be defined independently. This special case, suitably iterated, turns out to be enough; suitably iterated the special case *is* the general case.



### II. ANALYSIS

Before introducing the special case, a word about motivation. A philosophical account of X ought if possible to bring out de jure relations between X and other notions. It is not enough if X covaries as a matter of fact with ABC, even if the de facto covariation is counterfactually robust. (I include here metaphysically necessary covariation as the limiting case.) How well has this condition—the "de jure condition," let's call it—on a successful analysis been respected in the literature on intrinsicness?

This paper repays an old debt to Sydney Shoemaker. Years ago, as editor of *Philosophical Review*, he allowed me to cite a manuscript on intrinsicness that I was unable to provide at the time. He asked for it again in 1994, when he taught a seminar on intrinsicness; all I could offer was a one-page sketch of the main idea. Now you've got it, Sydney! I had help from Sally Haslanger, Ned Hall, Rae Langton, Jennifer McKittrick, Gideon Rosen, Alan Sidelle, Ted Sider, Judy Thomson, and Ralph Wedgwood. I wish I could acknowledge a greater debt to Peter Vallentyne's "Intrinsic Properties Defined"; it would have saved me a lot of trouble to have seen it earlier. An implicit subtheme here is that there is more to Vallentyne's approach than Langton and Lewis give him credit for.

<sup>1</sup> Jaegwon Kim, "Psychophysical Supervenience," *Philosophical Studies* 41 (1982): 51–70.



The first proposal worth mentioning is due to Jaegwon Kim. G is intrinsic according to Kim iff it is compatible with *loneliness*: the property of being unaccompanied by any (wholly distinct, contingently existing) thing.<sup>2</sup> It seems of the essence of intrinsicness that an intrinsic property should be possessable in the absence of other things; and so Kim's account does fairly well with the *de jure* condition. The problem with the account is that, as Lewis noticed back in 1983, it gives the wrong results.<sup>3</sup> Loneliness is as extrinsic as anything; but since it is a property compatible with loneliness (!), Kim would have to call it intrinsic.

What about Lewis's own account, according to which G is intrinsic iff given any  $x$  and  $y$  with the same natural properties,  $x$  is G iff  $y$  is G?<sup>4</sup> This gives more accurate results, but an element of *de facto*-ness has now intruded. If some natural property H should fail to be intrinsic, the account will overgenerate; it will call H "intrinsic" regardless. You may say that such a situation will never arise, and you may be right. The problem is that its never arising seems no more than a lucky accident.<sup>5</sup> There is no *in principle* reason why theorists of the quantum domain should not find themselves forced by nonlocality phenomena to count certain extrinsic properties as "ground floor" and natural. Because there is nothing in the nature of intrinsicness to prevent this, it is a matter of luck that Lewis's story succeeds as well as it does.

A modified account developed with Rae Langton draws only on a particular aspect of naturalness, viz., nondisjunctiveness.<sup>6</sup> According to Langton and Lewis, G is "basic intrinsic" iff (i) G and its negation are independent of loneliness and accompaniment and (ii) G and its negation are nondisjunctive. The intrinsic properties are the ones that never distinguish between things with the same basic intrinsic properties.

If you think that nondisjunctiveness comes closer than naturalness to being in the same conceptual ballpark as intrinsicness, then the modified account is more *de jure* than its predecessor. But an element of *de facto*-ness remains. Consider the property R of being the one and only round thing. R satisfies (i) and the first half of (ii); if it avoids being (basic) intrinsic, then, that is because it is the negation of the disjunctive property S of being non-round or else round and accompanied by a round thing.

Is S disjunctive, though? That it can be expressed in disjunctive terms doesn't count for much; roundness too can be expressed that way, e.g., as the disjunction of round-and-charged with round-and-not-charged. Better evidence

<sup>2</sup> Jaegwon Kim, "Psychophysical Supervenience," *Philosophical Studies* 41 (1982): 51–70.

<sup>3</sup> David Lewis, "Extrinsic Properties," *Philosophical Studies* 44 (1983a): 197–200.

<sup>4</sup> "New Work for a Theory of Universals," *Australasian Journal of Philosophy* 61 (1983b): 343–77. The formulation in the text leaves out some irrelevant complications.

<sup>5</sup> "Accident" is meant to be compatible with a fact's holding necessarily. If God and the null set are both necessary beings, then it is impossible for either to exist without the other. Still the correlation is accidental in the sense I have in mind.

<sup>6</sup> David Lewis and Rae Langton, "Defining 'Intrinsic'," *Philosophy and Phenomenological Research* 58 (1998).



of disjunctiveness would be a finding that S's "disjuncts"—the property of not being round, and that of being round and accompanied by something round—were perfectly natural. But they clearly are not, and so Langton and Lewis are led to maintain that S is at least much *less* natural than its "disjuncts." This means in particular that

S = accompanied by something round *if* round oneself  
is much less natural than

T = accompanied by something round *and* round oneself.

It may well be. But there is something uncomfortable about taking an intrinsicness-fact that is very clearcut—the fact that R is not intrinsic—and putting it at the mercy of something as controversial, and (apparently) irrelevant, as the relative naturalness of S and T. One feels that it ought not to *matter* to the issue of R's intrinsicness where S and T come out on the naturalness scale.

Of course, we don't want to be in the position of asking too much. Who is to say that intrinsicness bears *any* worthwhile de jure relations to other notions? Or maybe it bears them only to notions less fundamental than itself— notions that ought to be explained in terms of intrinsicness rather than the other way around.

The answer to this is that intrinsicness *does* appear to line up in nonaccidental ways with something quite fundamental: the relation of part to whole. It seems, for instance, as de jure as anything that

if *u* is part of *v*, then *u* cannot change intrinsically without *v* changing intrinsically as well.

And that

if *u* is part of *v*, then *u* and *v* have a region of intrinsic match.

And that

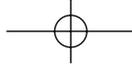
if *u* is properly part of *v*, then *u* and *v* have intrinsic differences.

So the materials for a more principled account are not obviously lacking. What remains to be seen is whether we can parlay one-way conditionals like the above into a de jure biconditional with intrinsicness all by itself on the left hand side.

### III. EXPANSION

The last of our three conditionals says that if one thing is properly part of another, then the two are intrinsically unlike. If you believe that, then you'll agree with me that *one* good way to arrange for intrinsic variation in the world outside of *x* is to *expand it*: add on something new. This immediately gives a necessary condition on intrinsicness:





- (\*) G is intrinsic only if necessarily, whether a thing is or is not G cannot be changed by adding a part to its containing world.

I say that the necessary condition is also sufficient. A counterexample would be a G that could be made to vary through intrinsic variation in the outside world, but not through the particular *sort* of intrinsic variation contemplated here: expansion. But for reasons about to be explained, such a G is not possible.

#### IV. WHY EXPANSION IS ENOUGH

Remember the intuition that we started with: If G is an extrinsic property, then it can be lost (gained) through intrinsic variation in the outside world. It follows from this that there are  $w$  and  $w'$  such that  $x$  has G in  $w$  but not  $w'$ , and the changes rung on  $w$  to get  $w'$  are *outside* of  $x$ .

Because the part of  $w$  that lies within  $x$  is left absolutely untouched,  $w$  and  $w'$  have a part in common: a part extending at least to  $x$ 's boundaries and possibly beyond. Focus now on this shared part. It could have existed all by itself; why not? Another way to put it is that there is a self-standing world (call it  $w''$ ) consisting of the shared part and nothing else.

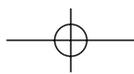
The question is, does G obtain in  $w''$  or does it not? However you answer, there are worlds  $u$  and  $v$  such that (i)  $x$  is part of  $u$  is part of  $v$ , and (ii) G is either lost or gained as you move from  $u$  to  $v$ . Suppose first that G does obtain in  $w''$ ; then G is lost as we move from  $u = w''$  to  $v = w'$ . If, on the other hand, G doesn't obtain in  $w''$ , then G is gained as we move from  $u = w''$  to  $v = w$ . Either way, G is sensitive to positive mereological variation in the outside world. Condition (\*) is not just necessary for intrinsicness but a sufficient condition as well.

#### V. ASSUMPTIONS

A few implicit assumptions must now be brought to light. But let me stress: *most of them are for simplicity only* and will eventually be given up. This applies in particular to assumptions (1)–(3), which together go under the name of “modal realism.” Later on (see the Appendix), we'll be abandoning possible worlds altogether. But for the time being our outlook is very close to that of David Lewis. We accept:

- (1) *Pluralism*: This actual world is only one of a large number of possible worlds. It is possible that BLAH iff in at least one of these worlds, BLAH is the case.
- (2) *Possibilism*: These other worlds exist—they are part of what there is—but they are no part of actuality.





- (3) *Concretism*: Other possible worlds are entities of the same basic sort as this world. They are maximal concrete objects that just happen not to be the same maximal concrete object that we around here inhabit.
- (4) *Mereologism*: There is only one part/whole relation worth taking seriously, a relation answering to the axioms of mereology.
- (5) *Recombinationism*: The space of worlds is closed under arbitrarily reshuffling of world-parts. As Lewis puts it, “patching together parts of different possible worlds yields another possible world.”<sup>7</sup>

Finally, a seeming corollary of (5),

- (6) *Inclusionism*: Some worlds contain others as proper parts.

I said that our outlook was “close to” that of Lewis; I might have said that we are going to be more Lewisian than Lewis himself. For it turns out that Recombinationism is *not*, the quoted passage notwithstanding, a view that Lewis accepts. (“An attached head does not reappear as a separated head in some other world, because it does not reappear at all in any other world.”<sup>8</sup>) Nor does Lewis accept Inclusionism. The reason is the same in both cases. Both assumptions represent worlds as having parts in common; and the sharing of parts is, according to Lewis, strictly forbidden. (A section of *Plurality of Worlds* is called “Against Overlap.” What part of “Against” don’t you understand?)

## VI. ACCIDENTAL INTRINSICS

If one looks, however, at what Lewis actually says in “Against Overlap,” his position turns out to be not quite as adamant as suggested:

my main problem is not with the overlap itself. Things do have shared parts, as in the case of the Siamese twins’ hand . . . what I do find problematic—inconsistent, not to mince words—is the way the common part of the two worlds is supposed to have different properties in the one world and in the other.<sup>9</sup>

Reading on, we learn what the inconsistency is. According to friends of overlap, Humphrey, who is part of this world and here has five fingers on the left hand, is also part of some other world and there has six fingers on his left hand. . . . He himself—one and the same and altogether self-identical—has five fingers on the left hand, and he has not five but six. How can this be? You might as well say that the shared hand of the Siamese twins has five fingers as Ted’s left hand, but it has six fingers as Ned’s right hand! That is double-talk and contradiction. Here is the hand. Never mind what else it is part of. How many fingers does it have?<sup>10</sup>

<sup>7</sup> David Lewis, *On the Plurality of Worlds* (London: Blackwell, 1986), 87–88.

<sup>8</sup> *Ibid.*, 88.      <sup>9</sup> *Ibid.*, 199.

<sup>10</sup> *Ibid.*, 199–200. This is not the first occurrence of “How many fingers?” in world literature; O’Brien puts the question to Winston in 1984.



Lewis calls this the *problem of accidental intrinsic*s. What matters to us is that it is a problem raised by some cases of overlap, but not all.

The problem does not arise, Lewis says, for Humphrey's *essential* properties,<sup>11</sup> however intrinsic they may be. No explanation is needed of how Humphrey can have different essential properties in different worlds, because he doesn't.

"Neither [does] it arise for Humphrey's extrinsic properties, however accidental." A thing's extrinsic properties are, implicitly, relations it bears to its surroundings. And there is no contradiction whatever in bearing a relation to one set of surroundings (being shorter than anyone in those surroundings, say) that you do not bear to another.

## VII. NOT ALL OVERLAP IS BAD

As far as the main argument of "Against Overlap" goes, a part of one world *can* recur in another world *provided that its intrinsic properties do not vary between the two worlds*.

One place we might seek assurances on this score is in the nature of the shared item. If the item was of a sort that possessed its intrinsic properties *essentially*, there would be no danger of its changing in intrinsic respects between the one world and the other.

A second place assurances might be sought is in the nature of the *relation* between the two worlds. Even if the item did not retain its intrinsic properties across *all* worlds, it might stay intrinsically the same across certain particular *pairs* of worlds—such as, for example, those standing in the part/whole relation. If so, there would be no objection to saying that the item was shared by those particular pairs of worlds.

And now our cup overfloweth; for our interest here is in a *case* of overlap—the case where one of two worlds is included in the other—that eludes Lewis's strictures in *both* of the ways just mentioned.

It eludes them in the first way because worlds are generally understood to have all of their intrinsic properties essentially.<sup>12</sup> World-inclusion avoids the problem of accidental intrinsic because it is *world-inclusion*.

Also, though, it avoids the problem because it is *world-inclusion*. One world is not part of another unless it recurs in it with the same intrinsic properties it had qua self-standing world.<sup>13</sup> Example: a world just like this one up to five minutes

<sup>11</sup> Meaning the ones he has in every world that contains him.

<sup>12</sup> Some might say that worlds cannot vary extrinsically either. But our resistance to extrinsic variation is much weaker, or the branching conception of worlds would not be so popular. See John Bigelow, "The World Essence," *Dialogue* 29 (1990): 205–17, for the view that a world's *intrinsic* nature is largely accidental to it.

<sup>13</sup> Advocates of the branching conception appear to take this for granted. A counterfactual world *w* branches from the actual world @ at time *t* iff world *w*, which has the history of @ up to time *t*



ago, when irresponsible atom-smashing experiments occurred causing everything to pop out of existence *now*,<sup>14</sup> is *not* an initial segment of actuality—whereas a world just like this one up to now, when everything *miraculously* pops out of existence, *is* an initial segment of actuality.<sup>15</sup>

### VIII. DEFINITION

I said that G is intrinsic iff, supposing that  $x$  has (lacks) G in  $w$ , G cannot be canceled (introduced) by moving to  $w'$ , where  $w'$  is  $w$  with the addition of a new part.

This may seem unnecessarily cagey. Why “canceled”? Why not just say that there can be no  $w'$  of the indicated type in which G fails to be exemplified by the original object  $x$ : the object that *did* exemplify G back in  $w$ ?

One worry would be that  $x$  might not *persist* into  $w'$ . But if we stick to our policy of departing from Lewis’s metaphysics only when absolutely necessary, this eventuality can be ruled out.

Remember, Lewis is a *mereologist*: he maintains that there is only one containment relation worth taking seriously, a relation (partially) characterized by the axioms of mereology.<sup>16</sup>

How does mereology eliminate the danger of  $x$  not surviving the trip to  $w'$ ? The story so far is that  $w$  and  $w'$  have a part in common that contains  $x$ . If containment is mereological part/whole, this means that  $x$  is a mereological part of the mereological overlap between  $w$  and  $w'$ . But the overlap between  $w$  and  $w'$  is (trivially) part of  $w'$ . So we can conclude by the transitivity of part/whole that  $x$  is part of  $w'$ , i.e., that  $w'$  contains  $x$ .

Now, if  $x$  persists into  $w'$ , then it seems clear that  $x$  is the thing in  $w'$  that G had better still attach to, if G wants to be regarded as intrinsic.<sup>17</sup> This gives us our first official definition of intrinsicness (‘ $<$ ’ stands for mereological inclusion):

as its entire history, is an initial segment of  $w'$ . That the relevant intrinsic features of @ “carry over” into  $w$  and thence into  $w'$  goes without saying.

<sup>14</sup> To go by a recent report in *Scientific American*, this is not altogether out of the question. Atom-smashers have felt moved on occasion to calculate the chances of their experiment flipping the universe into “null-state.”

<sup>15</sup> Lewis has other objections to overlap that are meant to apply even when there is no variation in intrinsic properties. He says, for instance, that overlap in the form of branching “conflicts with our ordinary presupposition that we have a single future” (*Plurality of Worlds*, 207). But these other objections are not advanced as decisive, and if overlap opens the door to a new account of intrinsicness, that would be an argument on the other side.

<sup>16</sup> It’s unique, but unselective (like identity). *Subset* is a type of part/whole for Lewis.

<sup>17</sup> Actually, perhaps it is not so clear. If  $x$  sits at a boundary of  $w$ , it could happen that  $x$  has a different shape in the larger  $w'$  than it had in  $x$ . Concessions about to be made in response to a different worry should alleviate this one as well.



- (1)  $G$  is *intrinsic* iff:  
for all  $x < w < w'$ ,  $x$  has  $G$  in  $w$  iff  $x$  has  $G$  in  $w'$ .

Roundness is intrinsic, by this definition, because for any  $w$  including  $x$  and any elaboration  $w'$  of  $w$ , if it is round in either, it is round in the other. Accompaniment is extrinsic because  $w$  and its elaboration  $w'$  can overlap on  $x$  while differing in the amount of company they provide it.<sup>18</sup>

## IX. ESSENTIAL EXTRINSICS

A consequence of (1) may give us pause. Call a property *absolutely essential* iff anything that has (lacks) it at all has (lacks) it in every world where it exists. Then *every absolutely essential property  $G$  is (1)-intrinsic*.<sup>19</sup> (If  $x$  has/lacks  $G$  essentially, there are not going to be *any* worlds in which  $x$  lacks/has  $G$ ; so there are not going to be any that contain  $w$  as a part.) This result gives us pause because it falls afoul of well-known examples of essential properties due to Kripke and others.

*Example 1:* According to Kripke, I am essentially descended from a certain zygote  $Z$ ; and it seems plausible as well that nothing can descend (in the relevant sense) from  $Z$  without being me. So *descending from  $Z$*  is an absolutely essential property. But on almost anybody's account, the zygote stopped existing before I started—whence *descending from  $Z$*  is an *extrinsic* property of mine. So how can (1) be believed when it tells us that extrinsicness and absolute essentiality are not compatible?

*Example 2:* Anything with the property of *being human* has that property essentially; and whatever lacks the property lacks it essentially as well. So *being human* is absolutely essential. But it is not intrinsic; how could it be when it involves an evolutionary lineage reaching back before the time not only of particular humans but of humanity itself? So *being human* is extrinsic and absolutely essential, again contrary to (1).

*Example 3:* Consider the property of *being me*, that is, *being SY*. Part of what it takes to be  $SY$  is to have the other two properties mentioned; so if they are

<sup>18</sup> Does (1) make a material object's location intrinsic to it? I don't think so. The problem arises only if one is a substantialist about space-time; it's only if location is the property of occupying such and such space-time points that  $x$ 's location can be expected to "follow it" from  $w$  into the more inclusive  $w'$ . But for  $G$  to be intrinsic, it's not enough that it can't be lost through world-expansion; a further requirement is that it can't be gained that way either. And this gives us a way out. No one supposes that the space-time points a table occupies are to be counted among the table's *parts*. So the table should be able to survive intact into a world  $v$  from which those points had been removed. Consider now  $v$  and the more inclusive world  $v'$  with which we began. That the table *gains* the property of occupying such and such space-time points in the transition from  $v$  to  $v'$  tells us that location is not intrinsic according to the definition in the text. (See also n. 29 below.)

<sup>19</sup> In this paper, " $x$  has  $G$  essentially" means that  $x$  has  $G$  in every world where it exists. The door is left open to a counterpart-theoretic rejoinder to the argument of the main text.

extrinsic, *being SY* is too. But we know from the necessity of identity (and distinctness) that *being SY* is absolutely essential. So *being SY* is an extrinsic and absolutely essential property, in defiance of (1).

Where are we? Lewis, recall, had his problem of *accidental intrinsics*—the problem being that accidental intrinsic properties are, initial appearances to the contrary, not possible. He “solved” this problem by caving in to it—*x*’s intrinsic properties do indeed have to be essential in one sense: *x* has them in every world where it exists—and then reestablishing his intended subject matter elsewhere—intrinsic properties do *not* have to be essential in the alternative sense that *x* shares them with all its counterparts.

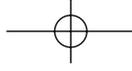
Our problem is something like the opposite of Lewis’s. It has to do not with accidental intrinsics but (*absolutely*) *essential extrinsics*. The problem is that (*absolutely*) essential extrinsic properties, despite making clear intuitive sense, are threatening to come out impossible. Our solution, in the next section but one, will have a similar shape to Lewis’s. First, there’ll be the caving in: if we limit ourselves to entities of such and such a type, then, yes, absolutely essential properties do have to be intrinsic. (But that is no paradox because entities of the type in question *should* have intrinsic essences.) Second, there’ll be the reconstitution elsewhere of our intended subject matter: if we let in entities of such and such *other* types, then absolutely essential extrinsics become once again possible.

## X. ALL-NOT-OVERLAP IS BAD

I just gave *being SY* as an example of an essential property that was extrinsic to its bearer. I didn’t say that *all* identity properties—all properties of the form *being x*—had to be extrinsic. Such a claim would not be plausible, and it has rarely been defended in philosophy.<sup>20</sup> If anything, the tendency has been to fall into the opposite error: the error of seeing *being x* as the intrinsic property par excellence. The right thing to say, with that error now exposed, is that some identity properties are intrinsic and others are not.

Take for instance a pair of protons, or two space-time points, or two identically shaped hunks of gunk. Do these have their (separate) identities intrinsically, or not? One can, of course, conceive of the first hunk being the hunk it is due, say, to its causal origins, which were different from the origins of the second hunk. But it takes imagination. The natural (and so presumably not absurd) thought is that the hunks are what they are as a purely intrinsic matter; each gets its identity from its gunk, and the gunk in hunk 1 is not numerically the same as the gunk in hunk 2. A venerable tradition even maintains that when you dig down to the

<sup>20</sup> But see Graeme Forbes, *Metaphysics of Modality* (Oxford: Clarendon Press, 1985), and my review in *Journal of Philosophy* 85 (1988), 329–337.



ultimate building blocks of reality—the level of protons and the like, or perhaps a deeper level still—identity is *always* intrinsic. One may agree with the tradition or one may not. But intrinsic identity is certainly not an *incoherent* idea, and so we should take care to allow for it in our theory of intrinsicness.

Allowing for it will be difficult, if overlap is forbidden. The quickest way to see this is via a widely accepted principle of Lewis’s connecting intrinsicness to duplication:

- (#) the intrinsic properties are the ones that never distinguish duplicates—including, crucially, duplicates in different worlds.

Suppose that overlap is prohibited. Then the property of *being x* distinguishes *x* from its otherworldly duplicates; if the duplicates were also *x* then we’d have overlap. It follows that whatever *x* may be, it has its identity extrinsically.

This is what I am calling problematic. A theory of intrinsicness should not predict right out of the starting gate that there is no such thing as intrinsic identity. But it will predict this if it accepts (#) and rejects overlap.

One response to the difficulty, proposed by Langton and Lewis, is to treat the theory as rendering no judgment in cases like this; the theory defines intrinsicness for “pure” or “qualitative” properties, not “impure” properties like *being x*.

But we might have had hopes of a unitary treatment of intrinsicness that applied to pure and impure properties alike.<sup>21</sup> If so, then our choices are either to reject (#) or to reconcile ourselves to the possibility of overlap. There is nothing wrong with (#), so we need to make our peace with overlap.<sup>22</sup>

The goal is a unitary treatment of intrinsicness—a theory as comfortable with impure intrinsics like *being x* as it is with pure intrinsics. How are we to tell if our attempt at a unified theory has succeeded? It would help if we had examples of intrinsic-natured items to test it against. Let’s adopt for purposes of this paper the view that worlds are made up of intrinsic-natured atoms *a*, *b*, *c*, . . .<sup>23</sup> And let’s make it a condition of adequacy on our (eventual) theory that it recognize *being a* as intrinsic for each atom *a*.

## XI. ESSENTIALISM AND MEREOLOGY

Back now to the issue we were dealing with before the digression. Analysis (1) may or may not succeed with accidental extrinsics; but to (absolutely) *essential*

<sup>21</sup> Lloyd Humberstone calls this intrinsicness-as-interiority. See his “Intrinsic/Extrinsic,” *Synthese* 108 (1996): 205–67.

<sup>22</sup> This is made easier by the fact that the principal objection to it doesn’t apply to the kind of overlap we are contemplating.

<sup>23</sup> Being composed of atoms  $a_1 \dots a_n$  will come out intrinsic on our account. See Humberstone, “Intrinsic/Extrinsic,” on the distinction between intrinsicness and nonrelationality.



extrinsics, like *being human*, it is quite blind. All such properties are found by the analysis to be intrinsic.

Could it be that we have been pitching (1) to the wrong audience? A case can be made that the Kripke examples do not refute (1) so much as calling attention to one of its presuppositions. (1) is an analysis for *monolithic mereologists*—people who see no real competitors in the part/whole department to the relation of *mereological parthood*.<sup>24</sup> And the problem of essential extrinsics *does not arise* for the monolithic mereologist, because for her *there are no such properties*.

Why not? Among the laws of mereology is one called the *unique-sum principle*: take any objects  $x_1, x_2, x_3, \dots$  you like, there is a unique object  $S(x_i)$ —their *sum*—with all the  $x_i$ s as parts and all of whose parts overlap the  $x_i$ s. If we accept this principle, then it becomes plausible to hold that an absolutely essential property has got to be intrinsic. What is the essence of the  $x_i$ s' unique sum going to be, after all, if not

- (a) to exist in exactly the worlds containing each  $x_i$
- (b) to exist at exactly the space-time positions occupied by any  $x_i$
- (c) to have the  $x_i$ s as parts, and
- (d) whatever is necessitated by (a)–(c)?<sup>25</sup>

Kit Fine calls the sum defined by (a)–(d) the *aggregate* of the  $x_i$ s, and remarks that “it has . . . often been supposed that aggregation is the only legitimate method of [summation].”<sup>26</sup> Whoever supposes that sums are aggregates and adopts our assumption above that every  $x$  is the sum, ultimately, of intrinsic-natured atoms, is well on the way to supposing that  $x$ 's essential properties are one and all intrinsic to it. For they will find it hard to resist the following line of argument:

Suppose that P is essential to  $x$ , where  $x$  is the sum of intrinsic-natured atoms  $x_i$ . Then P is necessitated by Q = the property of existing when and where the  $x_i$ s do, with them as parts. Given the intrinsicness of *being*  $x_i$ ,  $x$  possesses Q regardless of what goes on outside its boundaries. Hence it possesses P the same way; hence  $x$  has P intrinsically. Now, if a property essential to  $x$  is intrinsic to it, then a property that can *only* be had essentially (an absolutely essential property) can only be had intrinsically. A property that can only be had intrinsically, however, is an intrinsic property.<sup>27</sup> So every absolutely essential property is intrinsic.

<sup>24</sup> So-called because it is defined in large part by the fact that it satisfies the laws of mereology.

<sup>25</sup> Kit Fine, “Compounds and Aggregates,” *Noûs* 28 (1994): 137–58. Compare also Judy Thomson's “some-fusions,” in Thomson, “The Statue and the Clay,” *Noûs* 32 (1998): 149–73.

<sup>26</sup> Fine, “Compounds and Aggregates,” 138.

<sup>27</sup> The relation between intrinsic *properties* (our topic here) and intrinsic *predication* (having a property intrinsically) is nicely elaborated by Humberstone.



The unique-sum principle thus sits very well with (1); for it gives us a basis on which to *reject* the absolutely essential, extrinsic, properties that were causing (1) so much trouble.

If you are with me this far, though, you will see that the unique-sum principle sits very *ill* with the Kripkean perspective. From that perspective it is going to seem unduly limiting to suppose that the essence of an  $x$  made up of  $x_1, x_2, \dots$  has got to be (a)–(d) above.

If we're going to have a sum  $S(x_i)$  that exists in a world  $w$  iff *all* the  $x_i$ s exist there, why not a sum  $S^1(x_i)$  that exists iff *any* of the  $x_i$ s exist? Why not a sum  $S^2(x_i)$  whose existence in  $w$  requires only that *most* of the  $x_i$ s exist? Why not a sum  $S^3(x_i)$  which exists iff all and *only* the  $x_i$ s exist?<sup>28</sup> Why for that matter should there not be *lots of* sums  $S^\alpha(x_i)$ , alike in “ordinary” respects but differing in which of their properties they have essentially? And why should not some of these sums have extrinsic properties in their essences—as indeed one of the sums just mentioned already does?

These questions have no good answers, I think, or at least none that the Kripkean can be expected to find convincing. This is why I say that the moral of the Kripke examples is not that (1) is wrong, but that Kripkeans are not monolithic mereologists and so not among its intended audience.

## XII. INTRINSICNESS FOR THE REST OF US

Even if (1) can be defended as capturing *a* notion of intrinsicness—acceptable perhaps to monolithic mereologists—it remains to be seen how intrinsicness is to be defined for “the rest of us,” in particular, those who are convinced by the Kripke examples that the properties essential to a thing need not be intrinsic to it, or intrinsic at all.

Suppose that  $x$  has an extrinsic property  $G$  essentially. Then purely mereological changes in the part of  $w$  outside of  $x$  may well have the effect of removing  $x$  from the scene. So  $x$  will not be there to “witness”  $G$ 's extrinsicness in  $w$ . How then is  $G$ 's extrinsicness to be brought out?

The answer has got to be that it is not  $x$  but a *related* object  $x'$  whose  $G$ -ness in  $w'$  concerns us. An example will show what I mean. Suppose for argument's sake that creationists have got it partly right. The world did indeed spring into existence in 4004 B.C., complete with fossils, archeological remains, and other traps for the unwary. But, and this is where creationists overreach themselves, there was no act of God involved; the transition from nothing to something was a complete and utter ex nihilo miracle.

All of that given, it seems to me (and for the sake of the example, please let it seem to you, too) that it is essential to the kind *rabbit* that it originated

<sup>28</sup> That is, a sum  $S^3(x_i)$  that exists iff all the  $x_i$ s exist, and nothing disjoint from them exists.



more or less spontaneously, and, in particular, that it did not evolve from earlier kinds. It follows that Floppy here, who is essentially a rabbit, is essentially of a non-evolved kind. *Being of a non-evolved kind* is extrinsic, though; so the fact that Floppy has it essentially means that the property of *being identical to Floppy* is extrinsic as well.

Now consider a world  $w'$  that prefixes to  $w$  a long and complicated evolutionary history;  $w'$  is the sort of world that Darwinian types (wrongly) *believe* themselves to inhabit. When we look at the candidates for Floppy-hood in  $w'$ , we find that every one of them sits at the terminus of a continuous and biologically plausible developmental path tracking back to earlier species. Since it is essential to Floppy *not* to sit at such a terminus, Floppy does not exist in  $w'$ . What is it, then, that bears witness in  $w'$  to the extrinsicness of *being Floppy*? It has to be the “Darwinized” creature Floppy' that takes her place there.

Our challenge is to say how in general this  $x'$ —the object whose failure to exemplify  $G$  in  $w'$  reveals  $G$  as extrinsic—is to be identified. The outlines of the answer seem fairly clear:  $x'$  should be an object occupying exactly the piece or portion of  $w'$  that  $x$  occupied in  $w$ . But that still doesn't clue us in to the identity of the object  $x'$  that we want.

An advantage of the monolithically mereological approach was that it always made sense to speak of *the* occupant of the- $x$ -portion-of- $w$ -transplanted-into- $w'$ . The disadvantage, of course, was that the occupant was  $x$  itself, leading to the unhappy (from a Kripkean vantage point) results just noted: results that led us to reconsider our attachment to monolithic mereologism.

But although mereology can be pushed too far, it stands to reason that we should try to hold on to as much of it as we can, compatibly with the Kripkean data we are trying to accommodate. Why not let the view be that worlds are mereological *at bottom*—at the level of their “stuff” or “matter”—with the non-mereological aspects superimposed? Then we can say that at the bottom level, where mereology reigns, essences are always intrinsic; while at the higher Kripkean levels, where any sort of property can be essential, mereology graciously steps aside. There may be various ways of implementing this divide-and-conquer strategy, but the following seven step plan looks attractive:

*First*,  $a$ -sums:

The  $x_i$ s have a number of sums  $S^a(x_i)$ , exactly alike except in their essences, or (what comes to the same) their transworld careers.

*Two*, aggregates:

One of these  $a$ -sums is the *aggregate*  $S^0(x_i)$ ; its essence is to exist iff each of the  $x_i$ s do, when and where any of the  $x_i$ s do.

*Three*, pieces:

$x$  is a *piece* of  $y$  ( $x <^0 y$ ) iff there are things  $z_i$  such that  $y$  is the aggregate of all of them, and  $x$  is the aggregate of some of them.

*Four*, portions:

The *stuff* of a world  $w$  is the aggregate of all its atoms. Pieces of  $w$ -stuff, that is, aggregates of some or all of  $w$ 's atoms, are called *portions* of  $w$ .

*Five*, parts:

$x$  is a *part* of  $y$  ( $x < y$ ) iff there are  $a$  and  $z_i$  such that  $y$  is the  $a$ -sum of all of the  $z_i$ s, and  $x$  is the  $a$ -sum of some of them.<sup>29</sup>

*Six*, coincidence:

$x$  and  $y$  *coincide* iff they have exactly the same parts—equivalently, iff  $x$  is part of  $y$  and vice versa.

*Seven*, existence:

To *exist* in a world  $w$  is to be (a) a portion of  $w$ , or (b) constituted by (strictly: coincident with) a portion of  $w$ .

As the seventh step suggests, coincidence is the notion we need, but the definitions read better when framed in terms of constitution. A word should thus be said about what it is for  $x$  to constitute  $y$ , and why it does little harm to substitute constitution for coincidence. The account I have in mind is a slight variant of one recently proposed by Judith Thomson.<sup>30</sup> The idea is that  $x$ , although possessed of the same parts as  $y$ , “hugs” those parts more closely than  $y$  does:

$x$  constitutes  $y$  iff

- (a)  $x$  coincides with  $y$ ,
- (b) any part of  $x$  essential to it has parts that are not essential to  $y$ , and
- (c) no part of  $y$  essential to it fails to have parts that are essential to  $x$ .<sup>31</sup>

Because pieces of world-stuff hug their parts so very closely—*all* their parts are (modulo the coincidence relation) essential to them—almost anything coinciding with one is also constituted by it.<sup>32</sup> This is why not much harm can come of writing “ $x$  constitutes  $y$ ” instead of “ $x$  is coincident with  $y$ ” in the definitions to follow. When push comes to shove, however, it is always coincidence we really have in mind.

<sup>29</sup> Special cases aside,  $x$  is a part of  $y$  iff it occupies a subset of the space-time positions occupied by  $y$ . The special cases include, for instance, a bundle of compresent tropes vis à vis some particular trope  $t$  in the bundle; the bundle may occupy the same space-time position(s) as  $t$ , but it isn't part of  $t$ . Another example might be Casper the ghost passing through a mountain or a larger ghost. The special cases matter. An account of intrinsicness ought to distinguish the trope's intrinsic properties from those of the co-located bundle; likewise Casper and the co-located mountain-part. Intrinsicness is not a spatiotemporal notion except per accidens—the accident being that part/whole tends to be spatiotemporal.

<sup>30</sup> Thomson, “The Statue and the Clay.”

<sup>31</sup> At the risk of oversimplifying, the difference between this definition and Thomson's is that she has “some” at the beginning of (b) where I have “any.” This has the result, which I find unwelcome and she does not, that if Lump1 constitutes Goliath, then take any  $z$  you like (e.g., the planet Saturn), the fusion of Lump1 with  $z$  constitutes the fusion of Goliath with  $z$ .

<sup>32</sup> For an example of something coincident with a piece  $x$  of world-stuff in  $w$  that  $x$  does *not* constitute in  $w$ , consider a  $y$  that is “just like”  $x$  in  $w$  but exists in no other worlds. *Everything* about  $y$  is essential to it, so condition (b) cannot be met.



## XIII. INTRINSICNESS AND CONSTITUTION

All this time we have been grappling with a problem raised long ago for (1). The problem was that  $x$  may or may not carry over into  $w'$ . And if  $x$  does not exist in  $w'$ , then it makes no sense to ask whether  $x$  is G in  $w$  iff  $x$  is G in  $w'$ . Our proposed solution was to say that the portion of  $w$  that constitutes  $x$  there—the  $x$ -portion of  $w$ —*does* carry over into  $w'$ , where it constitutes an  $x'$  that *can* be assessed for G-ness in  $w'$ . If we modify (\*) to take account of this solution, we get

- (\*\*) G is intrinsic iff:  
 for all  $x < w < w'$ ,  $x$  has G in  $w$  iff  $x'$  has G in  $w'$ —where  $x'$  is whatever is constituted in  $w'$  by the  $x$ -portion of  $w$ .

An intrinsic property, in other words, is one that never distinguishes between  $x$  in  $w$ , on the one hand, and what  $x$ 's constituting matter goes on to constitute in expansions of  $w$ , on the other. Since constitution is itself explained in terms of part/whole, this brings us *close* to the sought-after reduction of intrinsicness to part/whole and modality.

## XIV. A FORK IN THE ROAD

But we are not there yet. The “whatever” in (\*\*) hides a choice. There are going to be *lots* of things  $x'$  constituted in  $w'$  by the  $x$ -portion of  $w$ . Which of them have to be G for a verdict of intrinsicness?

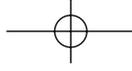
The answers that come to mind are: *at least one* of them has to be G, and *each* of them has to be G. So there are two disambiguations of (\*\*) to consider. These will be easier to read if we abbreviate “ $x'$  is constituted in  $w'$  by the  $x$ -portion of  $w$ ” to “ $x'$  is a copy of  $x$ ”:

- (2) G is intrinsic iff  
 for all  $x < w < w'$ ,  $x$  has G in  $w$  iff *some* copy  $x'$  of  $x$  has G in  $w'$ .  
 (3) G is intrinsic iff  
 for all  $x < w < w'$ ,  $x$  has G in  $w$  iff *every* copy  $x'$  of  $x$  has G in  $w'$ .

Before investigating how (2) and (3) differ, we should notice a way in which both of them improve on (1). They agree in allowing extrinsic properties to be essential.

*Example:* Suppose that  $w$  is the 4004 B.C. world described in section XII, imagined again as actual. And let  $w'$  be the more inclusive world that we wrongly *believe* to be actual. Our friend Floppy in  $w$  essentially possesses the property G of *not* sitting at the terminus of a long evolutionary history. But none of its copies in  $w'$  shares this property. It makes no difference, then, whether we go





along with (2) in demanding that all of Floppy's copies be G, or with (3) in demanding that some of them be G. Either way, the demand is not going to be met. And so both definitions classify G as extrinsic.

## XV. CATEGORICALS AND HYPOTHETICALS

Before trying to tease (2) and (3) apart, we need to fill in the background metaphysical picture some more, especially the all-important notion of coincidence. Coincidence theorists come in many shapes and sizes, but the relation I have in mind works as follows.

Imagine that we have before us the objects coincident in a world  $w$  with a portion  $p$  of  $w$ . These objects are, *categorically* speaking, just alike. To the extent that categorical properties fix perceptual appearance, they look just the same.<sup>33</sup> The differences between them are *modal*—one has essentially a property that another has only accidentally—or (to be a little more accurate) *hypothetical*—a matter of what they are like counterfactually, dispositionally, causally, or along any dimension that respects their behavior in other possible worlds.<sup>34</sup> By “casting our gaze” over these other worlds in imagination, we can “see” that the objects have different modal careers.

If this is how we understand coincidence, then (2) and (3) agree on the intrinsicness-status of categoricals. For let G be categorical. Then either *all* of the objects coincident in  $w'$  with  $x$ 's  $w$ -stuff are G, or *none* of them is G. And so the hypothesis on the right-hand side of (2) is equivalent to the one on the right-hand side of (3). The question is, how do (2) and (3) compare in their classification of non-categoricals?

## XVI. GENERIC VS. SPECIFIC INTRINSICNESS

Intrinsicness and categoricity are, of course, not the same. But there does seem to be a rather striking analogy between them. One can bring the analogy out like this: G is intrinsic iff whenever  $x$  has G in a world  $w$ , it has it in a manner insensitive to goings-on outside of  $x$ ; and it is categorical iff whenever  $x$  has G in a world  $w$ , it has it in a manner insensitive to goings-on outside of  $w$ .

<sup>33</sup> The compound of the  $x$ 's may well be categorically different from their aggregate. It will often be shorter-lived, for instance, since it goes out of existence when any of the  $x$ 's ceases to exist. This shows that Fine's compounding is not a sum operation in my sense (likewise Thomson's operation of all-fusion).

<sup>34</sup> I am not using “categorical” and “hypothetical” to help define “intrinsic.” I am using them to work out the consequences of definitions (2) and (3) for people who, like me, think that coincidents (things with the same parts) are categorically alike and distinguished by their hypothetical properties.



Someone who takes this analogy very seriously—someone who hears “outside of  $w$ ” as standing to “outside of  $x$ ” roughly as “outside of the galaxy” stands to “outside of this office”—will be tempted to conclude that sensitivity to circumstances outside of  $w$  is ipso facto sensitivity to circumstances outside of  $x$ . They will be tempted to conclude, in other words, that *hypothetical properties can never be intrinsic*. Call the notion of intrinsicness on which this is so—on which intrinsics have to be categorical—the *generic* notion. (“Generic” because it lumps other worlds in with other places as sources of extrinsicness.)

Whether they arrive at it by the indicated route or not, it is the generic notion that writers on intrinsicness often have in mind. Peter Vallentyne, for instance, says that “water-solubility and the like might seem like intrinsic properties, but once one recognizes [their] dependence on what the laws of nature are, it seems more correct to classify such properties as extrinsic.”<sup>35</sup> After all, Vallentyne argues, the laws of nature “are part of ‘the rest of the world’”—they are sensitive to changes in the world outside the object. If a lump of sugar can lose its solubility just by being placed in different surroundings, then solubility is not intrinsic.

Even if we agree, as not everyone does, that the laws in force at a given location depend on goings-on in “the rest of the world,” Vallentyne’s argument can be resisted. Dispositions are dispositions to behave in particular ways *in particular circumstances*. But if laws are circumstantial in the way that Vallentyne thinks, then there is no obvious reason why the circumstances should not be understood to include them.

An extrinsic disposition  $D_1$  can almost always be made into a (more) intrinsic one  $D_2$  by taking the external factors which make  $D_1$  come and go, and loading them into  $D_2$ ’s triggering circumstances. This is one way—a very old-fashioned way, to be sure—of understanding the difference between weight and gravitational mass. An object’s weight, on the old-fashioned account, is its disposition to depress a properly constructed scale so as to elicit a reading of so many pounds *in the local gravitational field, weak or strong as that field might be*. This nicely explains why weight is extrinsic: the reason your weight changes when you go to the moon, even though intrinsically you remain just the same, is that a different field becomes local. An object’s mass, on the other hand—its disposition to elicit a reading of so many pounds *in a gravitational field of such and such strength, wherever a field of that strength might be*—is not extrinsic, or anyway not *as* extrinsic as its weight.

All of this is to suggest that the generic notion of intrinsicness is not the only one we understand, or the only one we have need of.<sup>36</sup> Sydney Shoemaker in a celebrated paper distinguishes two types of causal power:

<sup>35</sup> Peter Vallentyne, “Intrinsic Properties Defined,” *Philosophical Studies* 88 (1997): 209–19.

<sup>36</sup> Some will see a different problem with the generic notion. They will object that the generic notion, far from being the *only* legitimate one, is in fact *illegitimate*, because dispositional properties are always intrinsic. See, e.g., David Lewis in “Finkish Dispositions,” *Philosophical Quarterly* 47 (1991): 142–58 and George Molnar, “Are Dispositions Reducible?” *Philosophical Quarterly* 49

A particular key on my key chain has the power of opening locks of a certain design. It also has the power of opening my front door. It could lose the former power only by undergoing what we would regard as a real change, for example, a change in its shape. But it could lose the latter without undergoing such a change; it could do so in virtue of the lock on my door being replaced by one of a different design. Let us say that the former is an intrinsic power and the latter is a mere-Cambridge power.<sup>37</sup>

“Mere-Cambridge” powers are in later writings called “extrinsic” powers.<sup>38</sup> The point either way is the same: if the key’s power to open a lock of such and such a type is intrinsic to it, then clearly we need a version of intrinsicness on which it does not entail categoricity. Since the reverse entailment does not hold in anyone’s book, we want intrinsic/extrinsic to crosscut categorical/hypothetical in every possible way.

Another example to consider here is identity. Atoms have their identities as an intrinsic matter. But *being x* is always hypothetical, since one can imagine a distinct thing  $x^*$  that is indiscernible from  $x$  when we bracket their counterfactual careers ( $x^*$  might differ from  $x$  just in the fact that it exists in fewer worlds).<sup>39</sup> The property of *being a*, where  $a$  is an atom, is thus intrinsic and hypothetical.

Other examples could be mentioned, but they aren’t really needed: it is obvious that we have, in addition to a generic notion of intrinsicness that entails categoricity, a *specific* notion that leaves an intrinsic property’s status as categorical undecided. The specific notion is the more discriminating and hence (it seems to me) the more useful one. A generically intrinsic property is just a specifically intrinsic one that happens also to be categorical; defining specific intrinsicness in terms of generic would not be so easy.

### XVII. INTRINSICNESS<sub>3</sub> = GENERIC INTRINSICNESS

Suppose that the space of particulars is *full*: for any assignment  $F$  of (coinstantiated) categorical properties to worlds, there is something  $x_F$  existing in just the world in  $F$ ’s domain and exemplifying in each of those worlds the properties that  $F$  assigns to it. Then, since a property is hypothetical iff it can be toggled by manipulating otherworldly categorical profiles, (3) makes all hypothetical properties extrinsic.

Why does (3) have this result? Setting  $w'$  in (3) equal to  $w$ , we see that  $G$  is intrinsic in the sense of (3) only if it never distinguishes coincidents in the same

(1999): 1–17. Both the always-extrinsic and the always-intrinsic positions are rejected here as too extreme.

<sup>37</sup> Shoemaker, *Identity, Cause, and Mind* (London: Cambridge University Press, 1984), 221.

<sup>38</sup> e.g., at 105–6 of *The First-Person Perspective and Other Essays* (London: Cambridge University Press, 1996).

<sup>39</sup> See my “Identity, Essence, and Indiscernibility,” *Journal of Philosophy* 84 (1987), 293–314.



world. No hypothetical property can meet this condition. If H is hypothetical, then there is a world  $w$  containing an  $x$  such that (i)  $x$  is H in  $w$ , but (ii) an  $x'$  differing from  $x$  only in its categorical properties in other worlds would not be H in  $w$ . By fullness, such an  $x'$  is bound to exist. So for any hypothetical property H, H's hypotheticality is "witnessed" by a pair of H-discernible coincidents in some world  $w$ . It follows that H is not intrinsic.

*Example:* Imagine a statue Goliath composed of a hunk of wax Lump1. Goliath is essentially of a certain intrinsic shape—it is essentially, let us say, so-shaped. But to be so-shaped is only accidental to Lump1, which might never have been formed into a statue at all. Since Goliath is coincident with Lump1, it follows that Goliath in  $w$  is a copy of something (Lump1) in  $w$  that differs from Goliath in point of essential so-shapedness. So the property of being essentially so-shaped is not intrinsic.

#### XVIII. INTRINSICNESS<sub>2</sub> = SPECIFIC INTRINSICNESS

Are hypothetical properties extrinsic in the sense given by (2)? Sometimes yes, sometimes no. Consider again the property of being essentially so-shaped. The reason this came out extrinsic<sub>3</sub> was that not everything coincident with Goliath in  $w'$  was essentially so-shaped; Lump1, for one, was so-shaped only accidentally.

But all that (2) asks is that *something* coincident with Goliath be essentially so-shaped. And given fullness, this condition is met for any world  $w'$  containing  $p$  = the Goliath-portion of  $w$ . For consider a super-Goliath stipulated to possess exactly the categorical properties of  $p$  in  $w'$  and to exist in no other worlds. Super-Goliath is an essentially so-shaped coincident of  $p$ . And that's precisely what is needed for essential so-shapedness to come out intrinsic<sub>2</sub>.

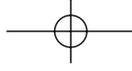
#### XIX. DOING IT WITH DUPLICATES

A property is intrinsic, according to Lewis, iff it never distinguishes duplicates. This reduces the problem of defining intrinsicness to the problem of explaining duplication. Suppose that we like this reduction, but don't like Lewis's explanation of what it is for  $x$  and  $y$  to be duplicates.<sup>40</sup> If we could obtain duplication from part/whole, that would be a move in the right direction.

Above we defined  $x'$  in  $w'$  to be a *copy* of  $x$  in  $w$  iff (a)  $w$  is part of  $w'$  (this to ensure that the  $x$ -portion of  $w$  persists into  $w'$ ), and (b)  $x'$  is constituted in

<sup>40</sup> Lewis relies on a distinction between natural and unnatural properties that he treats as primitive. One can agree that a primitive notion of naturalness is needed in philosophy, without agreeing that it is needed for the definition of intrinsicness, or that its availability is enough of a reason to use it.





$w'$  by the transplanted  $x$ -portion. The copying relation joined with its converse will be called *immediate duplication*. *Duplication* is the ancestral of immediate duplication: the relation that  $x_1$  in  $w_1$  bears to  $x_n$  in  $w_n$  iff there is a string of  $x_i$ s and  $w_i$ s ( $1 \leq i < n$ ) such that  $x_{i+1}$  in  $w_{i+1}$  is an immediate duplicate of  $x_i$  in  $w_i$ .

Now we are almost there; all that remains is to distinguish two different kinds of transworld similarity to serve the needs of our two definitions. Say that  $x$  in  $u$  and  $y$  in  $v$  *agree* on a property  $G$  iff:  $x$  has  $\pm G$  in  $u$  iff  $y$  has  $\pm G$  in  $v$ . And say that they *concur* on  $G$  iff:  $x$  has coincidents with  $\pm G$  in  $u$  iff  $y$  has coincidents with  $\pm G$  in  $v$ . Then the following are equivalent to (2) and (3):

(2')  $G$  is intrinsic iff duplicates always concur on  $G$ .

(3')  $G$  is intrinsic iff duplicates always agree on  $G$ .

Exactly as before, (2') gives us a modally neutral notion of intrinsicness—what above we called the specific notion—while (3') defines the generic notion whereby an intrinsic property has got to be “modally intrinsic” or categorical.

## XX. CONCLUSION

Lewis in “Extrinsic Properties” locates intrinsicness in “a tight little family of interdefinables,” and muses about the possibility of breaking out. Later, in “New Work for a Theory of Universals,” he sees his opening: reduce intrinsicness to duplication, and (the break-out point) duplication to naturalness. Still later, in “Defining ‘Intrinsic’,” he and Langton decide that they can make do with just an aspect or component of naturalness, viz., (relative) nondisjunctiveness.

But while that is certainly progress, neither of the proposed reducers—naturalness, nondisjunctiveness—*feels* like it has much to do with “what intrinsicness is.” Another way to put it is that both of the Lewis definitions trade on de facto connections with intrinsicness rather than de jure ones. It might be doubted, of course, whether intrinsicness *has* any interesting de jure connections with other notions. But conditionals like

(%)  $x$  is a part of  $y$  only if  $x$  cannot change intrinsically without  $y$  changing intrinsically as well.

show that this pessimism is unjustified. The de jure appearance of (%) led us to attempt a broadly mereological account of intrinsicness. Whether the account gives the right results in all cases, it does seem to lay the issue at a doorstep more in the right neighborhood.

The conditional (%), although putting us on the track of an account less accidental-feeling than Lewis’s, also suggests a way in which our approach might be thought to fall short of his. Naturalness has no chance whatever of being



explained in terms of intrinsicness; so a successful analysis of intrinsicness in terms of naturalness would probably amount to a reduction. One would truly have broken out of Lewis's "family of interdefinables."

An analysis in terms of part/whole, however, might or might not constitute a reduction, depending on whether part/whole was definable reverse-wise from intrinsicness. One is certainly accustomed to thinking of part/whole as the more basic of the two, but a look at (%) makes me wonder. Is a reversewise definition possible? If so, then we have at best enlarged Lewis's circle a little. If, on the other hand, part/whole deserves its reputation as primitive and undefinable, then perhaps we (too) can claim to have broken out.

## APPENDIX

This appendix explores how far we can disentangle ourselves from the modal realism presupposed in the main body of the paper. I'll suppose that we are confronted with a series of three objectors. ARG says that the *argument* we gave for our account of intrinsicness presupposed modal realism. PAL says that argument or no argument, the account is not *plausible* without modal realism. EXP maintains that our account is not even *expressible* without modal realism.

### ARG

A lot is riding on the idea that if one world is part of another, then this makes for an area of intrinsic match-up between them. The argument for this turned on Lewis's problem of accidental intrinsicness. But as Lewis would be the first to admit, accidental intrinsicness are a problem only for modal realists—only for those conceiving counterfactual worlds as big concrete objects on a par with the actual world. Modal realism is not a hugely popular doctrine.<sup>41</sup> How do you propose to defend your idea to the rest of us?

*Reply:* This misrepresents the argumentative strategy. Accidental intrinsicness came in not to *support* any claim about parts and intrinsicness, but to quiet a certain worry that the modal realist might feel. I said: look, to the extent that your opposition to overlap is based on the problem of accidental intrinsicness, you needn't be worried about overlap of the specific sort envisaged here. That one world can be part of another wasn't argued for at all; it was taken for granted. Also taken for granted was the claim that if one thing overlaps another, there is

<sup>41</sup> Not to mention that it has surprising consequences for intrinsicness. Assuming modal realism, accompaniment (as opposed to accompaniment by something you bear spatiotemporal relations to) comes out intrinsic; duplicates never differ with respect to it because *everything* is accompanied. Accompaniment has traditionally been the paradigm of an extrinsic property.

going to be complete intrinsic similarity with respect to the shared part. Nobody would question this in the intraworld case; why would the transworld case be different?

### PAL

I'll tell you why (it would be different in the transworld case). Not all transworld relations are created equal. On the one hand, you've got "genuine" transworld relations like *being the same color as*. These really do inherit the features of their intraworld originals. Transworld *same color as*, is transitive and entails *being colored*, just like the intraworld version.

On the other hand, you've got "degenerate" transworld relations like *touching*. All it can mean to say that  $x$  in  $w$  touches  $y$  in  $w'$  is that when you get them together in the same world, say  $v$ , you find that  $x$  touches  $y$  there. Degenerate transworld relations do *not* inherit the features of their intraworld originals; transworld touching, for example, is compatible with toucher and touchee each being quite alone in their original worlds.

Now if modal realism is correct, then just *maybe* there is a non-degenerate relation of transworld parthood. Otherwise though, parthood is like touching; when you try to apply it between worlds, it goes degenerate.<sup>42</sup> If that is right, then the fact that intraworld parts have a region of intrinsic match with their containing wholes says nothing whatever about the transworld case.

*Reply:* Transworld parthood is *not* degenerate. If it were, then (i) to be transworld part of  $y$ ,  $x$  would have to be intraworld part of  $y$  (in some salient world), and (ii) if  $x$  were intraworld part of  $y$  (in some salient world),  $x$  couldn't *avoid* being transworld part of  $y$ . I will argue that neither (i) nor (ii) is at all plausible.

Against (ii): The Kiwanis Picnic would have been considerably shorter had the world popped (miraculously!) out of existence just as the soda was being opened. Indeed, more is true: the picnic that would have been is a *proper part* of the picnic that actually was. According to (ii), though, this is impossible, for the picnic that was is a part of the picnic that would have been.

Why is that? Well, the picnic that would have been is none other than the picnic that actually took place; it is that picnic as it might have been rather than as it is. But if they are the same, then there is no question but that each intraworld includes the other; a thing always intraworld includes itself. By (ii), finally, this intraworld inclusion suffices for transworld inclusion. *The denier of genuine transworld parthood can thus make no sense of the idea that the picnic that would have been—the actual picnic with its tail end chopped off—is a proper part of the actual picnic.*

<sup>42</sup> For discussion, see Nathan Salmon's *Reference and Essence* (Princeton: Princeton University Press, 1981), 121ff.

Against (i): The Rainbow Rally includes what *would* have been the entire Spartacus League demonstration—would have been, if a few Spartacist rowdies wrongly supposed to be out of town had not turned up to demonstrate *against* the Rally. According to (i), the *would-be* demonstration is part of the Rally only if the *actual* demonstration is. So, although the would-be demonstration precisely *omits* the rowdies, it still has to pay the mereological price for their behavior. That seems absurd; *the Spartacist demonstration as it would have been is wholly included in the Rally.*

Degenerate transworld relations are a dime a dozen. Any “ordinary” relation R gives rise to a bunch of them by the formula: *x* in *w* bears transworld R<sup>*v*</sup> to *y* in *w'* iff *x* bears R to *y* in *v*. But these degenerate R<sup>*v*</sup>s are almost never what is meant by “transworld R.” It would be one thing if there were no alternative to the degenerate interpretation. But in this case there clearly is: it’s the interpretation we come to naturally when, e.g., we dispute with Lewis about whether there are counterfactual worlds with the actual world as a part. If “part” here stood for the degenerate relation, the answer would be obvious; it isn’t, so it doesn’t.

### EXP

Your account is committed to counterfactual worlds by virtue of explicitly quantifying over them. And the worlds you quantify over had better be Lewisian concrete worlds. Because it seems very doubtful that ersatz worlds, e.g., sets of propositions, are going to stand in the requisite inclusion-relations.

*Reply:* You’re right that the worlds of (2) and (3) are concrete. But just maybe the analysis can be (re)construed as speaking of worlds that, although they *could* have existed, as it happens do *not* exist—not even in a place bearing no spatiotemporal relations to the speaker.

The reason that the analysis appears to require existent concrete worlds is that it makes essential play with transworld relations such as transworld part/whole. How can the counterfactual picnic be (transworld) part of the actual one unless both are somehow *there*, in their containing worlds?

This is a fair question, but it has an answer; it had *better* have one, for colloquial English is thick with talk of transworld relations, and other worlds do not seem to come into the picture. You might say, for example, that

- (i) the car I would have had, if I’d installed afterburners in my old ’Vette, is faster than the Camaro I do have.

Does this commit you to the existence of a (counterfactual world containing a) counterfactual car, fitted out with afterburners as your actual car is not? Not at all. Your claim in essence is that

- (i’) it could have happened that I had a ’Vette that was faster than my Camaro is in actual fact.

And (i') makes no mention of counterfactual objects except in the scope of a modal operator. What remains to be seen is whether the same can be done with (2) and (3), that is, whether they too can be restated in a way that avoids wide-scope quantification over things that (modal realism aside) do not exist.

Before attempting this, let me concede right off that a formulation as colloquial as (i') is going to be hard to come up with. Our everyday modal devices are quickly pushed to their limits when the transworld comparisons become too involved, as in

- (ii) the car I would have had, if I'd installed afterburners in the 'Vette, is intermediate in speed between the one I would have had if I'd supercharged my Camaro and the Camaro as it is actually.

About the best we can do with this is

- (ii') it could have happened that, after supercharging, my Camaro was such that [it could have happened that I had a 'Vette with afterburners such that {the Camaro was faster than the 'Vette was and the 'Vette was faster than the Camaro *is*}]

This may not be wonderful English, but it seems close enough in *spirit* to (i') that it would be strange to discern here a qualitative change in subject matter—so that while (i') gave a partly modal description of actuality, (ii') described non-actual, merely possible, items, viz., other possible worlds and their inhabitants.

If that is right, then our next step should be to seek a Loglish-type regimentation of the language used in (i') and (ii'), in the hope that it permits a noncommittal reformulation of transworld-part talk, and ultimately a noncommittal reformulation of (2) and (3). A number of people have applied themselves to this sort of problem,<sup>43</sup> and they have had considerable success with a device called “multiple indexing.” Rather than trying to explain the idea from scratch, let me illustrate it with translations of our two target sentences. Think of “*c*” as a proper name of my actual Camaro, and “B” and “S” as standing for the relevant sort of afterburner-enhanced 'Vette and supercharged Camaro:

- (i'') possibly  $(\exists y) (\text{By} \ \& \ \text{actually}^c [y \text{ is faster than } c])$ .  
 (ii'') possibly<sub>1</sub>  $(\exists z) (z = c \ \& \ \text{Sz} \ \& \ \text{possibly}_2 (\exists y) (\text{By} \ \& \ \text{actually}_1^z \ \text{actually}_2^y \ \text{actually}^c [z \text{ is faster than } y \text{ is faster than } c]))$

Now let's try the method out on

- (iii) yesterday's Kiwanis picnic properly includes the picnic we would have had, had the world ended when the soda was being opened.

<sup>43</sup> See Forbes, *Metaphysics of Modality*, 89 ff. and references there.

Here it is, nearly enough, in noncommittal English:

- (iii') it could have happened that we had a picnic that was properly included in the picnic we did have.

And here it is in Loglish, with “*k*” standing for the picnic and “*P*” expressing picnic-hood:

- (iii'') possibly  $(\exists y) (\underline{P}y \ \& \ \text{actually}^k [y \text{ is a proper part of } k])$ .

*ital*

Now let's see what can be done with (2) and (3). I'll assume that, necessarily, there is one and only one world (which is not, of course, to say that “the world” is rigid). The letters “*u*” and “*v*” are world-variables, “*p*” and “*q*” range over world-portions, i.e., aggregates of atoms, “ $\approx$ ” expresses the coincidence relation, and “ $<$ ” stands for parthood:

- (3'') *G* is *intrinsic* iff:  
 necessarily<sub>1</sub>  $(\forall u) (\forall p < u) (\forall x \approx p) \text{ necessarily}_2 (\forall v) (\forall q < v) (\forall y \approx q)$   
 $(\text{actually}_1^{pu} \ \text{actually}_2^{qv} (p = q < u < v) [ \text{actually}_1 x Gx \leftrightarrow \text{actually}_2^y Gy ])$

✓ **superscript**

*G* is intrinsic<sub>3</sub>, in other words, when the following holds necessarily: if a mass of atoms that composes *x* would, had a more inclusive world obtained, have composed *y*, then *x* is *G* iff *y* would have been *G*. The definition of intrinsicness<sub>2</sub> is nearly the same, except that the two coincidence-quantifiers get moved to either side of the final biconditional:

Could (3'') and (2'') be compressed to eliminate the dangling last lines?

- (2'') *G* is *intrinsic* iff:  
 necessarily<sub>1</sub>  $(\forall u) (\forall p < u) \text{ necessarily}_2 (\forall v) (\forall q < v) (\text{actually}_1^{pu} \ \text{actually}_2^{qv} (p = q < u < v) \rightarrow [ \text{actually}_1 (\forall x \approx p) \pm Gx \leftrightarrow \text{actually}_2 (\forall y \approx q) \pm Gy ])$

⌋  
⌋  
⌋  
⌋

The  $\pm G$ s towards the end are to indicate a conjunction of two biconditionals, one with *G* unnegated on both sides, one with it negated. Translated into English, what (2'') says is that *G* is intrinsic<sub>2</sub> iff necessarily, any mass of atoms that composes only *G*s (non-*G*s) would still have composed only *G*s (non-*G*s), had a world obtained of which the actual world is only a part.

REFERENCES

Bigelow, John (1990). “The World Essence”. *Dialogue* 29: 205–17.  
 Fine, Kit (1994). “Compounds and Aggregates”. *Noûs* 28: 137–58.  
 Forbes, Graeme (1985). *Metaphysics of Modality*. Oxford: Clarendon Press.  
 Humberstone, Lloyd (1996). “Intrinsic/Extrinsic”. *Synthese* 108: 205–67.  
 Kim, Jaegwon (1982). “Psychophysical Supervenience”. *Philosophical Studies* 41: 51–70.  
 Lewis, David (1983a). “Extrinsic Properties”. *Philosophical Studies* 44: 197–200.

- Lewis, David (1983b). "New Work for a Theory of Universals". *Australasian Journal of Philosophy* 61: 343–77.
- (1986). *On the Plurality of Worlds*. London: Blackwell.
- (1991). "Finkish Dispositions". *Philosophical Quarterly* 47: 142–58.
- Lewis, David, and Langton, Rae (1998). "Defining 'Intrinsic' ". *Philosophy and Phenomenological Research* 58.
- Molnar, George (1999). "Are Dispositions Reducible?" *Philosophical Quarterly* 49: 1–17.
- Salmon, Nathan (1981). *Reference and Essence*. Princeton: Princeton University Press.
- Shoemaker, Sydney (1984). *Identity, Cause, and Mind*. London: Cambridge University Press.
- (1996). *The First-Person Perspective and Other Essays*. London: Cambridge University Press.
- Thomson, Judy (1998). "The Statue and the Clay". *Noûs* 32: 149–73.
- Vallentyne, Peter (1997). "Intrinsic Properties Defined". *Philosophical Studies* 88: 209–19.
- Yablo, Stephen (1987). "Identity, Essence, and Indiscernibility". *Journal of Philosophy* 84: 293–314 [Chapter 1 in this volume].

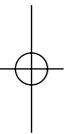


# 3

## Cause and Essence

### 1. COMMENSURATION

“What *kind* of thing is a cause, or an effect? And supposing that  $x$  and  $y$  are of the right kind, what should they be like *specifically*, for  $x$  actually to cause  $y$ ?” This paper considers both questions in the belief that their answers are connected. Among other things, it argues that causes and effects have *essences*; that causal properties are *hypothetical* rather than categorical; and that how a thing is essentially is *relevant* to its causal properties.<sup>1</sup> All of this is supposed to follow on a determined application of the principle that causes are (in a sense to be explained) ‘commensurate’ with their effects.



Of our two questions, the first, about the type of thing *apt* to stand in causal relations, is relatively recent.<sup>2</sup> Traditionally, causal theorists concentrated on the second, namely, what makes one thing of that type the cause of another? Hume and Mill, for example, address it in the form: Must the cause take in *everything* required for the occurrence of its effect, or can it comprise just a selection out of that material? In saying that “we must reject the distinction between *cause* and *occasion*, when suppos’d to signify anything essentially different from each other”,<sup>3</sup> Hume hints at an affirmative answer, an answer explicit in Mill’s thesis that the true cause is seldom “a single antecedent” of the effect, but rather “the sum of several antecedents; the concurrence of all of them being requisite to produce, that is, to be certain of being followed by [the

Something like the present approach to causation was proposed in the last two chapters of my dissertation (1986, ‘Things’, University of California, Berkeley). In Yablo (1987) the essentialist half of the story is laid out in some detail, and the connection with causation briefly indicated; this paper takes the cause/essence connection as its main object. I am grateful to Louise Antony, Paul Boghossian, Sin Yee Chan, Donald Davidson, John Drennan, Graeme Forbes, Sally Haslanger, Jaegwon Kim, Louis Loeb, Vann McGee, Sarah Patterson, Gideon Rosen, Larry Sklar, William Taschek, Ken Walton, and Crispin Wright for discussion and advice. Research for this paper was supported by the National Endowment for the Humanities and the Social Sciences and Humanities Research Council of Canada.

<sup>1</sup> Actually I will be arguing these points in connection with causes and their causal properties only, not effects and theirs. But most of what I have to say applies to effects *mutatis mutandis*.

<sup>2</sup> See Vendler (1962, 1967/1975), Davidson (1967/1980a), Kim (1973), Lewis (1986a), and Bennett (1988).

<sup>3</sup> Hume (1968; p. 171).



effect]”.<sup>4</sup> Both philosophers also ask whether the cause can include anything *not* needed for the effect. Again, they tend to agree that it cannot. Thus Hume:

In almost all kinds of causes there is a complication of circumstances of which some are essential, and others superfluous; some are absolutely requisite to the production of the effect, and others are conjoin'd only by accident.<sup>5</sup>

By means of his “rules by which to judge of causes and effects”, he says,

we learn to distinguish the accidental circumstances from the efficacious causes; and when we find that an effect can be produc'd without the concurrence of any particular circumstance, we conclude that that circumstance makes not a part of the efficacious cause, however frequently conjoined with it.<sup>6</sup>

Mill, of course, credits his famous “methods of experimental enquiry” with a similar capability. On the Hume/Mill theory, then, the cause includes all, and only, factors required for the effect’s occurrence.<sup>7</sup>

Probably Hume and Mill went too far in imposing so rigorous a condition on causes. Few would deny that my slamming the door startled the hamsters on the ground, e.g., that it was enough that the door got slammed, irrelevant that it was me who slammed it. But even if they were wrong in their specific thesis that the cause comprises all and only what the effect requires, they were surely onto a correct general principle: *nothing causes an effect that leaves out too many relevant factors, or brings in too many irrelevant ones*. True causes, as I will put it, are *commensurate* with their effects.

## 2. SIZE AND STRENGTH

Two related issues are tangled up in the idea of commensuration: one about the cause’s *size* or *extent* in space and time; and another about the cause’s reach

<sup>4</sup> Mill (1950, Bk. III, Ch. V. §3). True,

each and every condition of the phenomenon may be taken in its turn and, with equal propriety in common parlance . . . spoken of as if it were the entire cause. And, in practice, that particular condition is usually styled the cause whose share in the matter is superficially the most conspicuous, or whose requisiteness to the production of the effect we happen to be insisting on at the moment.

But “philosophically speaking”, the cause “is the sum total of the conditions . . . the whole of the contingencies of every description, which being realized, the consequent invariably follows” (loc cit.).

<sup>5</sup> Hume (1968, p. 148).

<sup>6</sup> Ibid., p. 149.

<sup>7</sup> This is at any rate what they say about general causes—sunlight as the cause of day—and there is nothing to suggest that they want to deal differently with the singular case—the lightning bolt as the cause of the stampede. Some of Mill’s examples are explicitly singular, e.g., when a man slips on a ladder his death is due not just to the fall but also “the circumstance of his weight” (Mill, loc. cit.). However the situation is complicated by the fact that neither author is very attentive to the distinction between singular and general causation.

along a quite separate axis. Take size first, concerning which Mill has a good example:

When the decision of a legislative assembly has been determined by the casting vote of the chairman, we sometimes say that this one person was the cause of all the effects which resulted from the enactment. Yet we do not really suppose that his single vote contributed more to the result than that of any other person who voted in the affirmative.<sup>8</sup>

Because votes cast *elsewhere*, and presumably *earlier*, contributed to the effect, the alleged cause is *extensively* incomplete. Obviously a complementary possibility is that it should be extensively excessive. According to Mill, what causes the coming of day is “the existence of the sun . . . and there being no opaque medium in a straight line between that body and the part of the earth where we are situated”, and this “without the addition of any superfluous circumstance”.<sup>9</sup> Since it *is* superfluous that night obtained earlier, the given conditions *plus* the fact of recent night are rejected as involving more than the effect needed.

That causes should arrange themselves along spatiotemporal lines is not surprising, nor is it surprising that they should be comparable in spatiotemporal extent. In Davidson’s well-known discussion of Mill’s strictures on causation, he introduces, inadvertently I think, a new dimension of comparison:

‘The cause of this match’s lighting is that it was struck. —Yes, but that was only *part* of the cause; it had to be a dry match, there had to be adequate oxygen in the atmosphere, it had to be struck hard enough, etc.’ We ought now to appreciate that the ‘Yes, but’ comment does not have the force we thought. It cannot be that the striking of this match was only part of the cause, for this match was in fact dry, in adequate oxygen, and the striking was hard enough.<sup>10</sup>

What I find troubling here is the absurdity of the misapprehension that Davidson’s bland reminders seem aimed at correcting, *viz.*, that what is partial about the striking *per se* is that it *lacks* the causally important properties mentioned. To think *that* would be to think that the striking *per se* was a lackadaisical striking of a wet match in a vacuum, or, even more incredibly, that it was somehow indeterminate on all these points.

<sup>8</sup> Mill, *loc. cit.* Assuming that the earlier voting did not influence his thinking, the chairman’s vote falls short of the true cause in what might be called *latitudinal* extent: extent along lines cross-cutting the lines of causal influence. Building on some enigmatic remarks of Hume, Russell (1963) poses an interesting problem of *longitudinal* extent. Normally we think of causes as taking time, the later portions depending on the earlier. But unless we are prepared to countenance the temporal equivalent of action at a distance, “it would seem that only the later parts can be relevant to the effect . . .” (p. 135). Apparently, then, the earlier portions must be written out as superfluous. An analogous argument shows that the only *real* causation is simultaneous causation! (See Hume 1968, pp. 76, 174–75; Ducasse 1969, pp. 44ff.; Taylor 1962–63/1975, p. 41; Lucas 1962, pp. 63–65; and Beauchamp and Rosenberg 1981, pp. 182ff.).

<sup>9</sup> Mill (1950, Bk. Ch. III. V, §6).

<sup>10</sup> Davidson (1967/1950a, pp. 155–56): “What is partial in the sentence, ‘The cause of the match’s lighting is that it was struck’”, he continues, “is the *description* of the cause . . .”.

Well, what else could the ‘yes, but’ comment be getting at? Another of Davidson’s examples has a bridge’s collapse said to be caused, not by the bolt’s snapping *as such*, but by its snapping so *suddenly*.<sup>11</sup> Taking inspiration from the match example, we might protest as follows:

How does the first snapping fall short of the second? From their descriptions it seems that there is to be a distinction in point of suddenness. But *both* are sudden (there is no *non-sudden* snapping here in question). So the problem is the same as before: to explain how things can differ on a property which, manifestly, they both possess.

Is this really so mysterious, though? If things both of which possess a property are to differ in point of that very property, the difference can only lie in the *manner* of its possession. To give this difference a name, only one of the two occurrences has the property *constitutively*. If the bolt’s suddenly snapping does better than its snapping *per se* at causing the bridge’s collapse, that is because it is *constitutively* sudden, whereas the other is sudden just as a matter of fact.

### 3. PROBLEMS FOR CONSTITUTIONALISM

With this we pass from a rather familiar position on our second question—that causes should be commensurate with their effects—to a rather unorthodox position on the first—that causes have some of their properties constitutively and others not. To repeat the steps: commensuration presupposes that causes are proportionable to their effects; for that, they must be comparable in size and strength; but to be comparable in strength, they must be of such a type as to show an inherent preference for certain of their properties over others. The view that they do show such a preference I will call *constitutionalism* about causes.

By comparison with the more usual view that causes are *concrete*, in the sense of possessing all of their properties on a par,<sup>12</sup> constitutionalism has a lot of

<sup>11</sup> Or at least that is how I would describe the case. Davidson speaks rather of the collapse’s being caused by the *fact* that it gave way so suddenly. Further the ‘caused’ here “is not the ‘caused’ of straightforward, singular causal statements, but is best expressed by the words ‘causally explains’”, the latter to be understood as a non-truth-functional connective (Davidson 1967/1980a, pp. 161–62). To appreciate why he thinks the example needs special treatment, we need to see what his objection is to the “straightforward” reading. The objection in the text is offered as in the spirit of his remarks elsewhere.

<sup>12</sup> For an influential early treatment of *concreta*, see Ducasse (1969, pp. 62ff. and *passim*). Davidson sometimes puts the concrete theory like this: concrete particulars are “endlessly redescrivable”. Presumably the idea is that entities not so redescrivable are made in the image of language, so “intensional” rather than concrete. Such a view harks back to Quine’s early criticism of quantified modal logic that it was committed to a realm of “dubious entities” insusceptible of analytically inequivalent description (Quine 1953, pp. 152ff.). As Quine subsequently realized, though, anything that can be talked about at all can be specified in analytically inequivalent ways (*loc. cit.*); hence no ontological distinction whatever is marked by the proposed condition. Neither is any

explaining to do. What are constitutions, that things with different of them can still be overwhelmingly similar in other respects? How can things as similar as *that* still differ in what they cause? And of what possible relevance can their constitutions be to their ability to influence events?

All of these are understandable concerns, but I suspect that it is the last that mainly accounts for constitutionalism's continuing unpopularity. Hume and Mill are again a good place to start. Both lay great stress on what I have been calling commensuration, sizewise and strengthwise, too.<sup>13</sup> Yet, although this *sounds* like a situation tailor-made for the constitutional approach, and although both occasionally 'talk the talk',<sup>14</sup> they seem in the end not to approve of the invidious distinctions that constitutionalism requires.<sup>15</sup> They had reasons from elsewhere in their philosophies for disliking such distinctions, of course, and questions of causal ontology were not in any case foremost in their minds.<sup>16</sup>

distinction marked by the redescrivability condition, for literally anything, intensions included, can be endlessly redescribed. Similarly unhelpful are the following remarks: "[It is wrong to think] that we have not specified the whole cause of an event when we have not wholly specified it" (Davidson 1967/1980a, p. 156); "[Not] every deletion from the description of an event represents something deleted from the event described" (op. cit., p. 157); "[A]n event is something . . . concrete with features beyond those we use to describe it" (Mackie 1974, p. 256); and "causes and effects are events in the sense of concrete occurrences exemplifying features over and above the ones we hit upon for describing them" (Beauchamp and Rosenberg 1981, p. 248). For again, entities of *every* sort admit of more or less informative description, and none can be described completely. (It was a mistake in any case to try to characterise the concrete by a contrast with the intensional. By a concrete ontology is meant one too coarse-grained to supply distinct entities for, e.g., 'the bolt's snapping' and 'the bolt's snapping so suddenly' to refer to. But Davidson nowhere argues, and it is not true, that the rejected distinction can be provided for only on an intensional ontology. Like Quine and others reared on the Church/Carnap interpretation of quantified modal logic, he tends not to recognise any middle ground between the concrete and the intensional, such as the essential could conceivably occupy. As a result his own formulations are apt to mislocate his position in the space of contemporary options, which is why I have preferred to characterise the concrete theory as in the text.)

<sup>13</sup> At least they give the appearance of admitting both types of commensuration. But again, interpretation is complicated by their willingness to run particular and general causes together. (Relevant texts are Hume (1968, pp. 148–49, 173–75), and Mill (1950, Bk. III, Chs. VI–X, *passim*). See also Hume's remark (1963, pp. 150–51) that "[i]f the cause, assigned for any effect, be not sufficient to produce it, we must either reject that cause, or add to it such qualities as will give it a just proportion to the effect". Here Hume seems to be making an epistemological point rather than a metaphysical one.)

<sup>14</sup> For example, Hume writes that "where several different objects produce the same effect, it must be by means of some quality, which we discover to be common amongst them. For as like effects imply like causes, we must always ascribe the causation to the circumstances, wherein we discover the resemblance" (1968, p. 174). This leads Davidson to speculate that "it is not events, but something more closely tied to the descriptions of events, that Hume holds to be causes" (Davidson 1967/1980a, p. 150). A less extreme reaction would be to say that Humean causes cannot, consistently with the passages in question, be regarded as *concrete* events. Kim (1973) and Beauchamp and Rosenberg (1981) argue in effect that Hume *should* have conceived his causes as constitutional events given the uses he has in mind for them.

<sup>15</sup> Mill in particular maintaining that "individuals have no essences" (1950, Bk. I, Ch. VI, 3).

<sup>16</sup> "[I]t should be clear not only that Hume did not address the question of the ontology of causation directly, but that no consistent theory about what kinds of items are causally related is likely to emerge solely from textual analysis" (Beauchamp and Rosenberg 1981, p. 249).

But another factor in their neglect of constitutionalism might have been simply this: that they did not see what *help* it would be in the quest for commensurate causes. For how can a thing's preferences among its properties affect its causal powers?

Whether precisely this difficulty occurred to Hume and Mill or not, in Davidson's critique of Mill it comes up often:

How could Smith's actual fall, with Smith weighing, as he did, twelve stone, be any more efficacious in killing him than Smith's actual fall?<sup>17</sup>

By emphasising Smith's weight we might improve our *explanation* of his death, but to think that the *cause* could be improved by a similar emphasis is just a confusion. Here is Jonathan Bennett in the same spirit: if "what got him down was not (so much) her refusing him but (more) her refusing him rudely", and if these "differ only in their constitutions, not in their characters", then

the refusal that did not get him down (so much) was just as rude as the other, but it lacked the other's depressive powers because rudeness was not in its constitution. . . . That, however, should make us suspicious. All the popular theories of event causation . . . agree with clamorous common sense that the causal powers of any event depend upon what it is like, what properties it has, what its character [as opposed to its constitution] is.<sup>18</sup>

To answer this would be to explain how a thing's constitution can be relevant to what it causes. But first we need to say something about constitution itself.

#### 4. ESSENCE<sup>19</sup>

By essentialism, I mean the view that things have some significant quota of their properties essentially, the rest only accidentally.<sup>20</sup> So understood, essentialism has a surprising consequence for identity: things exactly alike in every ordinary respect (location, shape, size, mass, microphysical makeup, etc.) may nevertheless fail to be numerically the same. That will be the case, whenever *x* and *y* agree in their ordinary properties but differ in which of those properties they possess essentially.

For the most part, essentialists concede this result,<sup>21</sup> but try to soften it by postulating intimate identity-*like* relations compatible with strict distinctness: composition, instantiation, generation, composita, and the like. But, not to

<sup>17</sup> Davidson (1967/1980a,b, p. 150; the example is Mill's).

<sup>18</sup> Bennett (1988, pp. 81–82). I should point out that I am wilfully misreading Bennett. His objection is to the causal relevance of constitution, not in the sense of essence, but in the sense of Kim's property-exemplification theory (that he is not himself puzzled about the causal relevance of constitution-*qua*-essence is clear from pages 54ff.). However his language can be read as expressing the more general concern raised in the text.

<sup>19</sup> Some of the material in this section is adapted from Yablo (1987).

<sup>20</sup> As usual, a thing's essential properties are those it could not have existed without.

<sup>21</sup> Lewis, interestingly enough, concedes it for events but not objects. See Lewis (1971; 1986a).

minimise these relations' importance, their differences distract us from a more fundamental relation they imply in common. Suppose we use the term *categorical* for properties whose possession by a thing  $x$  is a matter of  $x$ 's actual condition, as opposed to what it would or could have been like (other properties, e.g., dispositional and modal properties, are *hypothetical*).<sup>22</sup> Then the relation I am thinking of is this:  $x$  is *coincident* with  $y$  iff they have their categorical properties in common.

Beware of reading the categorical/hypothetical distinction as just a paraphrase of the accidental/essential distinction. For one thing, properties are accidental only in relation to specified particulars and worlds, but they are categorical *simpliciter*. More revealingly, for  $P$  to be accidental to  $x$  at  $w$  is partly a matter of how  $x$  is at  $w$  ( $x$  must have  $P$  at  $w$ ) and partly a matter of how it is at other worlds ( $x$  must lack  $P$  in at least one such). But,  $P$  is categorical if its attaching to  $x$  at a world is *wholly* a matter of how  $x$  is at that world, absolutely without regard to its otherworldly behaviour. Thus it should come as no surprise that hypothetical properties, for instance dispositions, can be accidental; and that categorical properties can be essential, as mountains are (I suppose) essentially spatially extended.

Though the distinctions are different, they can be related through a certain notion of *essence*. Essences I will understand as sets of essential properties:  $x$ 's essence is a set of properties essential to  $x$ ,  $y$ 's essence is a set of properties essential to  $y$ , and so on. But which of a thing's essential properties go into its essence?

The simplest proposal, obviously, would be to include *all* of them. Unfortunately, essences so defined will not meet our needs. What we are after, among other things, is an account of comparative strength as discussed in Section 2; and such an account will presumably be in terms of inclusion relations between essences. The problem is that these inclusion relations are liable to be disrupted, if essences are not somehow restricted. Allowing *identity with  $x$*  into  $x$ 's essence, for example, precludes the possibility of a  $y$  whose essence includes everything in  $x$ 's essence and more besides. And the effect of allowing  $x$ 's *kind* into its essence is to ruin the chances for a thing  $y$  whose essence exceeds  $x$ 's by properties which things of that kind possess at best accidentally.<sup>23</sup>

Is there an approach that avoids this difficulty? For the essences of nonidenticals to be comparable, they should be drawn from a pool of properties such that any particular such property's modal status—essential or accidental—is without undue prejudice to the modal status of the others. Since to include properties like these in essences does nothing to impede the later entry of their companions,

<sup>22</sup> Notice the parallel with the more familiar occurrent/nonoccurrent and intrinsic/extrinsic distinctions; categorical/hypothetical is to the modal dimension roughly as occurrent/nonoccurrent and intrinsic/extrinsic are to time and space.

<sup>23</sup> To events of the kind *stabbing*, for example, it is not essential that the victim subsequently dies. So if *being of the kind stabbing* was allowed into the essence of Brutus's stabbing Caesar, there would be no possibility of building up to Brutus's killing Caesar by adding in Caesar's subsequent death.



I call them *cumulative*. Although I do not know how to specify the cumulative properties outright, their most important features can be summed up in a simple condition. Letting  $x$ 's *essence* be the set of cumulative properties that it possesses essentially, and letting  $x^+$  *strengthen*  $x$  ( $x^+ \geq x$ ) if  $x$ 's essence is a subset of  $x^+$ 's, the condition is that

- (K) For all  $x$ , for all possible worlds  $w$ , for all sets  $S$  of cumulative properties:  $x$  exists in  $w$  and possesses there every member of  $S \Leftrightarrow$  there is an  $x^+ \geq x$  which exists in  $w$  and to which every member of  $S$  belongs essentially.

That is,  $x$  exists and possesses a set of cumulative properties (in a world) iff there exists also (in that world) a strengthening of  $x$  to which those properties attach essentially.

Applying (K) in the right-to-left direction, with  $S$  equal to the empty set, gives:

- (1) If  $x^+ \geq x$ , then necessarily, if  $x^+$  exists, so does  $x$ .

Applied from right to left, with  $S$  the difference between  $x^+$ 's essence and  $x$ 's, it implies:

- (2) If  $x^+ \geq x$ , then necessarily, if  $x^+$  exists,  $x$  possesses every property in  $x^+$ 's essence.

And we get:

- (3) If  $x^+ \geq x$ , then necessarily, if  $x$  exists and possesses every property in  $x^+$ 's essence, then  $x^+$  exists,

by (first) running (K) from left to right with  $S$  the essence of  $x^+$  — this to obtain the existence of an  $x^* \geq x^+$  — then (second) using (1) to infer the existence of  $x^+$  itself. To illustrate, if the speeding strengthens the driving, then the driving occurs in every world in which the speeding does; and in all such worlds, and only them, the driving is done at a high speed.

When one thing strengthens another, as the speeding strengthens the driving, the difference between them is *merely* hypothetical if any difference is (ultimately it comes down to the fact that  $x^+$  essentially possesses properties that are accidental to  $x$ ). But, if only hypothetical properties can distinguish  $x^+$  from  $x$ , then a property is categorical only if it *cannot* distinguish them:

- (C)  $P$  is categorical  $\Rightarrow$  for all  $x$  and  $x^+$  such that  $x^+ \geq x$ , and for all possible worlds  $w$  in which both exist,  $x$  has  $P$  in  $w$  iff  $x^+$  has  $P$  in  $w$ .

Thus  $x$  and its strengthening  $x^+$  are categorically indiscernible, or coincident, in every world where both exist:

- (4) If  $x^+ \geq x$ , then necessarily, if  $x$  and  $x^+$  exist, they are coincident.





For instance, the bolt's suddenly snapping is categorically indiscernible from its snapping per se, not just in this world but in every other where they exist together. Understanding a relation to hold *essentially* between  $x_1$  and  $x_2$  when necessarily, it holds if  $x_1$  and  $x_2$  exist, (4) can be put by saying that if one thing strengthens another, they are essentially coincident.

Strengthening is not the only form of coincidence, though, nor do all coincidence relations hold essentially. Imagine that  $x$  and  $y$ , although neither strengthens the other, have (in world  $w$ ) a strengthening  $z$  in common. Then by (4),  $z$  is coincident in  $w$  with  $x$  and  $y$ , whence  $x$  and  $y$  are coincident in  $w$ , too. To turn this observation to advantage, assume that:

- (U) For every  $x$ , and every world  $w$  in which  $x$  exists, there is an  $x_w \geq x$  which exists in  $w$  alone.

Again by (4),  $x$  and  $x_w$  have the same categorical properties in  $w$ ; and since  $x_w$  exists in a single world only,  $x_w$  possesses these properties essentially. Thus every detail of  $x$ 's worldly condition is essential to  $x_w$ , which licenses us in referring to it as  $x$ 's *state* in  $w$ . Now suppose that  $x$  and  $y$  are in the *same* state in  $w$ , that is, there is a  $z$  existing in  $w$  alone that strengthens both of them. Then by the same argument as before, they are coincident in  $w$ :

- (5) Necessarily, if  $x$  and  $y$  are in the same state, they are coincident.

With the help of one further assumption, we can strengthen (5) to a necessary biconditional:

- (6) Necessarily,  $x$  and  $y$  are in the same state iff they are existent and coincident.

That assumption, independently plausible, is that:

- (N)  $x$  and  $y$  are distinct  $\Rightarrow$  either they exist in different worlds, or they are noncoincident in some world where both exist.

Suppose that  $x$  and  $y$  exist in  $w$  and are coincident there. By (4) and (U),  $x_w$  and  $y_w$  are existent and coincident there, too. Since  $x_w$  and  $y_w$  exist in no other world but  $w$ , they exist in the same worlds and are coincident in all of them, which by (N) makes them identical. So  $x$  and  $y$  are in the same state in  $w$ . That proves (6)'s right-to-left direction; (1) and (5) imply the other. Assuming that things in the same state in one world can be in distinct states in others, (6) supports the claim above that it is possible for coincidence relations to hold accidentally. (An example might be the coincidence of a statue with its constituent clay.)<sup>24</sup>

<sup>24</sup> (N) has one other consequence worth noting:

- (7)  $x$ 's essence =  $y$ 's essence  $\Rightarrow x = y$ .

If  $x$  and  $y$  have the same essence, then each strengthens the other. By (1), they exist in the same worlds; by (4), they are coincident in each of these worlds; and now (N) implies that they are identical.



## 5. CONSTITUTION AS ESSENCE

At the end of Section 2, constitutions were proposed as a way of reconciling the following assumptions:

- (i) the bolt's suddenly snapping has different causal powers from its snapping *per se*;
- (ii) there must be some prior difference between them, to account for this;
- (iii) this prior difference must be in point of suddenness; but
- (iv) they are exactly alike in every ordinary respect, suddenness included.

These can all be true together, I said, if, although both snappings are *sudden*, only one of them is sudden *constitutively*. But I acknowledged that the proposal may seem to raise more questions than it answers. What are constitutions? How is it that disparately constituted entities can be otherwise so similar? And how can things as similar as *that* still differ in their causal powers?

What recommends essentialism is that it gives a way of approaching these questions. Constitutions are essences. Because essential properties, that is properties of the form *being essentially P*, are hypothetical, things can differ essentially while still being categorically just the same.<sup>25</sup> Lastly, if causal properties are hypothetical (see below), then it is only to be expected that categorical duplicates will sometimes differ in their effects.

To me this is motivation enough for the essentialist account, or at least for pursuing it further. But some may feel that it runs so far counter to modal intuition that it cannot be taken seriously as it stands (at best it bears some fortuitous formal analogy to the correct account).

Well, what *are* our intuitions here? Few would find it strange to say, of Brutus's stabbing Caesar, that Caesar might have survived it if the knife had been blunter; or of the sinking of the Titanic that, if certain hatches had held, it might have stretched out over hours or days.<sup>26</sup> Yet to say these things of Brutus's *killing* Caesar (that Caesar might have survived *it*) or of the Titanic's *swiftly* sinking (that *it* might have taken days) is incomprehensible.<sup>27</sup> These are

<sup>25</sup> Indeed some essential differences *entail* categorical indiscernibility (Section 4, (4)).

<sup>26</sup> I have given the descriptions wide scope to deflect the common charge that event essentialism owes all its plausibility to scope confusions (Davidson pp. 170–71; see also Neale 1990, pp. 145ff.). Thus it might be held that 'necessarily, Brutus's killing Caesar was not survived by Caesar' resembles 'necessarily, the U.S. President is an American citizen' in being defensible on the narrow scope reading only. But then 'regarding Brutus's killing Caesar, it could not have been survived by Caesar' should be just as preposterous as 'regarding the U.S. President, he could not have failed to be an American citizen'; which I submit it is not.

<sup>27</sup> Not that all, or only, the features by which we identify a cause or effect are essential to it. To turn a well-known remark of Davidson's to a foreign purpose, "we must distinguish firmly between causes and the features we hit on for describing them" (Davidson 1967/1980a, p. 155). Nobody

far from being isolated hunches. With expert medical attention Caesar might have pulled through; in that case, the stabbing would have occurred in the killing's absence. Likewise the Titanic's sinking, if had been sufficiently prolonged, would have occurred without the Titanic's swiftly sinking.<sup>28</sup> Essentialism may not be the only interpretation of these data but it is certainly the most straightforward one.

Essentialism about causes is a theory of their common nature; to stretch a phrase, it is a theory of "what they are". However it could be held—and I do hold<sup>29</sup>—that essentialism applies to all particulars whatever, regardless of categorial or other differences. So if the question is, not what causes *are*, but what they are *as opposed to other things*, essentialism is rather a minimal position. Nevertheless I propose to add very little more: only the commonplace, anticipated

would think, simply on the basis of their descriptions, that it was essential to the revolutionary upheaval recounted in *Ten Days That Shook the World* to have featured in that work, or accidental to Versailles' most famous postwar conference that it involved the European powers. How far a cause's essence can be judged from its description is a complicated matter; appearances can and do mislead. For example: at first 'the rabbi's noisy praying' and 'the rabbi's noisily praying' may seem coreferential. But then a puzzle arises. Much as to speak of the rabbi's blue prayer-book is to speak of his prayer-book, identifying it by its color, to speak of his noisy praying is to speak of his praying, identifying it by its volume. Accordingly there is no more reason to think of his noisy praying as *essentially* noisy than of his blue prayer-book as essentially blue. Yet we can make little sense of a situation in which his *noisily* praying, though it still occurs, is not noisy. So the descriptions are not coreferential after all. This should give some idea how unobvious the rules are linking a cause's description to its essence. Here are some extremely amateur hypotheses. One, in so-called *imperfect* nominals (see Bennett 1988), the converted verb typically indicates an essential property. Thus Amelia's *flying* to Marseille could not have been a swimming there. Two, whether *perfectly* nominalised verbs indicate essential properties depends on whether the verb's perfect form amounts to a genuine sortal. Flights are essentially flights; but at least some failures, or so we imagine, *could* have been successes. Three, recalling that perfect nominals are modified adjectivally, imperfect nominals adverbially, only the second sort of modification connotes essentiality. Thus the heavy fall Amelia took might have been lighter if she had managed to catch herself; her falling heavily would not have occurred at all.

<sup>28</sup> For more on nominalisation, see Vendler (1962; 1967/1975), Kim (1973), Chomsky (1975), Thomason (1985), Lewis (1986a), and Bennett (1988, chs. 1 and 2). Kim holds, as I do, that 'his praying' and 'his noisily praying' are non-coreferential. However Kim's neglect of the perfect/imperfect distinction leads him to class 'his noisy praying' with the latter rather than the former. Vendler and Bennett agree with me that 'his noisy praying' and 'his noisily praying' are non-coreferential, but only because they construe imperfect nominals as standing quite generally for a different *type* of entity than perfect—facts, rather than events. That goes too far. Facts have their reality by timelessly and placelessly *obtaining*, and the rabbi's noisily praying is, like his noisy praying, something that *happens* at a particular time and place. Moreover the rabbi's noisily praying was, like his noisy praying, *noisy* (or we standing outside the rabbi's door would not have heard it); but the *fact* that he prayed noisily was not noisy, and it is not what we heard. (Several authors have noticed that (i) 'the rabbi's noisy praying' and (ii) 'the rabbi's noisily praying' seem to be related roughly as (iii) 'the cat, which purrs' and (iv) 'the cat's purring'. To strengthen the analogy we might try postulating for 'NP, which VPs' and 'NP's VPing' a common transformational ancestor; say, 'NP + VP'. Applied to 'the cat + purr', the suggested transformations yield (iii) and (iv). Applied to 'his praying + occur in a noisy manner', they yield, not (i) and (ii) exactly, but the roughly equivalent (i\*) 'his praying, which occurs in a noisy manner' and (ii\*) 'his praying's occurring in a noisy manner'.)

<sup>29</sup> Yablo (1987).

in one or two incautious formulations along the way, that they are things which take place or happen, thus *events* or *occurrences* or *happenings* of some sort.

Notice that even this would be to say too much, if events were (as on some theories) inherently coarse-grained.<sup>30</sup> Whatever else is true of causes, there needs to be a *distinct* one for each of the finely discriminated causal roles they are called on to fill. An opposite worry would be that the proposal is not explicit enough; that the needed distinctions are so extraordinary that they can be provided for only on some special basis, e.g., by intensionalising events, or endowing them with internal structure.<sup>31</sup> But, on the first point, it is only *concrete* events that are inherently coarse-grained, and, on the second, fine-graining comes automatically with our freewheeling background essentialism. So the most straightforward course is to treat causes simply as events with essences; or, since *everything* has an essence, simply as events.<sup>32</sup>

## 6. CAUSAL PROPERTIES AS HYPOTHETICAL

Above we distinguished categorical properties from hypothetical, but said nothing about how the distinction bears on causal properties. To any ordinary way of thinking, I suggest, causal properties are hypothetical. For instance, we see, or think we do, a strong connection between  $x$ 's causing  $y$  and its being such that without it  $y$  would not have occurred. But whether  $x$  has the latter property depends on what goes on in nonactual situations.

Another sort of evidence that causal properties are hypothetical comes from our essentialism about causes. Recall that properties like that of causing a certain effect are liable to discriminate on grounds of strength, e.g., the bolt's suddenly snapping has different effects from its snapping *per se*. But events related by strengthening are categorically alike. If there is a causal difference between them, then, it can only be hypothetical.

So causal properties *must* be hypothetical to tell categorical duplicates apart; it is a further point that their being hypothetical clarifies *how* they do this, or, to

<sup>30</sup> Davidson (1967/1980a); Mackie (1974, ch. 10).

<sup>31</sup> See Vendler (1962, 1967/1975); Kim (1973); Gibbard (1975); Dretske (1977); and Bennett (1988).

<sup>32</sup> Perhaps it would be useful to compare the approach taken here with Kim's property-exemplification theory. That theory discovers a uniform object-property-time structure in events, and calls events identical iff they have the same constitutive elements. These identity-conditions being intraworld only, the theory draws no contrast whatever between an event's essential properties and its accidental ones. What our approaches have mainly in common, and their principal contrast with the concrete theory, is the insistence on fine distinctions between events that are in some attenuated sense 'the same'. However only essentialism has a story to tell about what these fine distinctions are—categorical indiscernibility tempered by hypothetical difference—and only essentialism can predict them on the basis of its analysis of the events themselves. For example, from the essences of the bolt's snapping *per se*, and its snapping so suddenly, it *follows* that they are coincident but hypothetically unlike. Given just the object-property-time analyses of these events, one has so far not even an *interpretation* of their 'sameness', much less an argument for it.

put it the other way around, how essence manages to be causally relevant. For a thing's essential properties enjoy a certain preeminence relative to its other hypothetical properties, causal properties included. To possess hypothetical property  $P$  is to lead a certain kind of counterfactual life, amounting finally to the possession of such-and-such categorical properties in such-and-such counterfactual situations. But, how a thing categorically comports itself across its counterfactual environments is a function of how it is essentially. Here then is the form (only that) of a mechanism connecting essence to causal powers. Sections 7–10 suggest one way, perhaps not the only way, in which the mechanism might actually work.

### 7. EFFECTS AS CONTINGENT ON THEIR CAUSES<sup>33</sup>

In the *Enquiry Concerning Human Understanding*, Hume describes cause and effect as items such that “if the first . . . had not been, the second never had existed”.<sup>34</sup> As an analysis or definition of causation, this is of course extremely doubtful.<sup>35</sup> But as the *de facto* generalisation that, other things equal,  $x$  causes  $y$  only if:

(C) If  $x$  had not occurred, then  $y$  would not have occurred,

Hume's remark verges on truism. Calling  $y$  *contingent* on  $x$  when  $x$  and  $y$  satisfy (C), the truism is that effects are, other things equal, contingent on their causes.<sup>36</sup>

<sup>33</sup> From this point on, I make a distinction between *necessary* and *essential* properties:  $x$ 's necessary properties are those it cannot exist without, and its essential properties are those in its essence (see Section 4). Similarly, I distinguish between *contingent* and *accidental* properties:  $x$ 's contingent properties are those it can exist without, and its accidental properties are those of its contingent properties which are eligible to belong to essences, i.e., the cumulative ones. So a property is essential (accidental) iff it is necessary (contingent) and also cumulative. Occasionally it may seem that I am treating a property  $P$  as essential whose cumulativity is doubtful. In all such cases I should be read as speaking rather of  $P^0$ , defined so that  $x$  has  $P^0$  in a world iff something coincident with  $x$  in that world has  $P$  there. (From its definition it follows that  $P^0$  is categorical, and on assumptions defended elsewhere, to be categorical is equivalent to being cumulative (Yablo 1987, prop. 3)). For example, I *do not* maintain that the essence of Brutus's killing Caesar includes the property  $P$  of causing Caesar's death. Surely in fact  $P$  is not a property of Brutus's killing Caesar at all; when someone dies as Caesar did, it is not the killing that kills him, but the stabbing. What *is* essential to the killing is to be *coincident* with something, in this case the stabbing, causative of Caesar's death. But this last is a categorical property, hence cumulative.

<sup>34</sup> Hume (1963, p. 83). This is not, of course, his preferred description of the causal relation; see Lewis (1973/1986b).

<sup>35</sup> Most of the counterexamples are to the condition's sufficiency for causation (Kim 1974). On the necessity side we have mainly the problem of causal preemption to deal with (Lewis 1973/1986b). Preemption happens when, although  $y$  results from  $x$ , if  $x$  had not occurred  $y$  would still have occurred as the result of some other cause. So  $x$  causes  $y$  but  $y$  is not contingent on it. (For reasons developed by Lewis, in some such cases  $x$  and  $y$  *do* stand in the ancestral of the contingency relation, i.e., they are connected by a chain of events each contingent on its predecessor. Thus contingency's ancestral comes closer to being necessary for causation than contingency itself. I ignore this refinement here.)

<sup>36</sup> See Lyon (1967) and Lewis (1973/1986b). All I can offer in defence of my resort to an admittedly fallible condition is that: (i) virtually *every* known condition is fallible; (ii) the condition

(Here and throughout, ‘if it had been that  $P$ , then it would have been that  $Q$ ’ is counted true in a world  $w$  iff  $Q$  is true in the  $P$ -world best resembling  $w$ .<sup>37</sup>)

Because the contingency condition makes no overt mention of essence, its essence-sensitivity can easily be overlooked. Suppose that it was irrelevant to Socrates’s death that he guzzled the hemlock, rather than simply drinking it. Then Xanthippe is mistaken when, disgusted at her husband’s sloppiness, she complains that his *guzzling* the hemlock caused his death. Assuming that the drinking would still have occurred if the guzzling hadn’t, contingency explains the error nicely. Even without the guzzling, the death would still have followed on the drinking (the details would naturally have been different). If not for the drinking, though, the death would not have occurred at all. So the effect is contingent on the weaker antecedent, but not the stronger.<sup>38</sup>

Implicit in the example is an argument that as properties irrelevant to  $y$  accumulate in  $x$ ’s essence,  $y$ ’s contingency on  $x$  is threatened. Suppose that  $x$  possesses many such irrelevancies essentially. Then  $x$ ’s absence from the nearest  $x$ -less world is liable to signify nothing more than the failure there of a property not implicated in  $y$ ’s production. Since the failure of *that* sort of property should not take  $y$  out of existence, it will be *false* that  $y$  would not have occurred if  $x$  had not, i.e., false that  $y$  is contingent on  $x$ . To the extent then that effects are contingent on their causes, it damages  $x$ ’s credentials for the role of cause if irrelevant properties are too often essential to it.<sup>39</sup>

## 8. CAUSES AS ADEQUATE FOR THEIR EFFECTS

Most counterfactual theories of causation put the contingency condition front and centre; they refine it in light of counterexamples and surround it with caveats, but genuinely collateral conditions are rare. To the outsider this

holds in *general*; and (iii) some *such* condition will presumably have a place in anyone’s counterfactual theory of causation.

<sup>37</sup> So we opt for the Stalnaker rather than the Lewis variant of the standard semantics—allowing, with Stalnaker, that where it is indefinite what the closest  $P$ -world is, this can make for indeterminacy in the counterfactual (Stalnaker 1981a, 1981b). Specifically: ‘if it had been that  $P$ , then it would have been that  $Q$ ’ is true (false) iff on every (no) admissible choice of closest  $P$ -world, the closest  $P$ -world is a  $Q$ -world. Might-counterfactuals ‘if it had been that  $P$ , then it might have been that  $Q$ ’ are true iff the corresponding would-counterfactuals ‘if it had been that  $P$ , then it would have been that not- $Q$ ’ are untrue, i.e., iff on at least one admissible choice of closest  $P$ -world, the closest  $P$ -world is a  $Q$ -world.

<sup>38</sup> Lewis (1986a) puts contingency to similar use.

<sup>39</sup> Here is the argument less metaphorically: as  $x$ ’s essence accumulates causally irrelevant properties, the chances increase that  $x$  is survived, in the nearest  $x$ -less world  $v$ , by a weakening  $x^-$  whose essence falls short of  $x$ ’s essence in causally irrelevant respects only. Since  $x^-$  preserves  $x$ ’s causally important properties on the scene,  $y$  should still occur in  $v$ , contrary to the contingency condition.

is surprising, since naively one expects some sort of adequacy condition complementary to contingency. Here are two reasons why such a condition seems desirable.

When Xanthippe attributed Socrates's death to his guzzling the hemlock, she overestimated the actual cause. But causes can also be essentially *underestimated*. Whatever Admiral Poindexter might think, it was not his testifying to Congress, *as such*, that caused his downfall; rather his *lying* to Congress was to blame. But where Xanthippe's mistake is subject to correction by the contingency condition, Poindexter's opposite error is not, for the indictment and so on were no less contingent on his testifying than on his lying. Or suppose someone suggests Zsa Zsa Gabor's *driving* (rather than her speeding) through the police radar as what led to her detention, or attributes the officer's abrasions to her *touching* his face (rather than her slapping it). Again, these attributions strike most of us as wrong, but the contingency condition is unbothered.

Apart from the examples, a new condition is needed to complete a powerfully, if obscurely, felt symmetry in the character of causation: if the cause is a *that without which not*, it is also a *that with which*.

Probably the main reason for adequacy's neglect is that this second notion has resisted all attempts at counterfactual analysis. Neither of the obvious candidates seems to work. When (C)'s antecedent and consequent are negated, we get:

(A<sub>1</sub>) If  $x$  had occurred, then  $y$  would have occurred also;

when we transpose them the result is:

(A<sub>2</sub>) If  $y$  hadn't occurred,  $x$  wouldn't have occurred either.

Although (A<sub>1</sub>) is not *wrong* as a condition on causation, it follows trivially from a more basic condition—that  $x$  and  $y$  should actually occur—to which adequacy is intuitively quite unrelated.<sup>40</sup> (A<sub>2</sub>)'s problem is worse: it approaches on being *incompatible* with a more basic condition and hence with the causal relation itself. To cause  $y$ ,  $x$  must be causally prior to it. But if  $x$  is causally prior to  $y$ , then it is probably *not* the case that it would not have occurred if  $y$  hadn't; rather  $x$  would have occurred as ever, but the causal train from  $x$  to  $y$  would have been derailed at the last minute.<sup>41</sup>

So where (A<sub>1</sub>) is vacuous, (A<sub>2</sub>) is for the most part unsatisfiable. Avoiding both extremes is the condition that:

(A) If  $x$  had not occurred, then *if it had*,  $y$  would have occurred as well<sup>42</sup>

<sup>40</sup> (A<sub>1</sub>) comes from Lewis (1973/1986b). Since it is Lewis, too, who notices that (A<sub>1</sub>) is trivial when  $x$  and  $y$  occur, I assume that he is not offering it as an interpretation of adequacy.

<sup>41</sup> (A<sub>2</sub>) is from Mackie (1974). See Lewis (1979/1986c, 1973/1986b) for the argument about "back-tracking" counterfactuals.

<sup>42</sup> Rasmussen (1982) contains the only explicit reference to condition (A) I have seen. There Rasmussen argues, fallaciously I think, that (A) follows from (C) on the hypothesis that both  $x$  and  $y$  occur.

(i.e.,  $y$  occurs in the nearest  $x$ -containing world  $u$  to the nearest  $x$ -omitting world  $v$  to actuality). (A) *would* be vacuous—it would follow automatically from the existence of  $x$  and  $y$ —if the nearest  $x$ -containing world  $u$  to the nearest  $x$ -omitting world  $v$  was, whenever  $x$  actually existed, the actual world. But why should it be? More likely, the actual world sits in the interior of a neighbourhood of  $x$ -containing worlds, whose outskirts contain worlds *nearer* to the nearest  $x$ -omitting world than the actual world is. Unlike (A<sub>1</sub>), then, (A) is not vacuous. Unlike (A<sub>2</sub>), it doesn't ask too much: it will be satisfied whenever it is correct to say 'suppose that  $x$  had not occurred; then  $y$  *would* have occurred if  $x$  had'. This seems, in any case, a reasonable test of intuitive adequacy. For the question is whether  $x$ , introduced into the actual circumstances *minus*  $x$ , brings  $y$  in its train. And it is hard to think what the actual circumstances *minus*  $x$  could be, if not the circumstances that *would* have obtained if  $x$  had not occurred.<sup>43</sup>

Imagine a bridge designed so that, given time to respond, it shifts its weight away from failing bolts. To take advantage of this design feature, special 'soft' bolts are used which snap readily but seldom abruptly. This particular day, alas, our bolt has just begun to give way when molecular bonds along the fracture line improbably deteriorate. The snapping is thus accelerated, and the bridge, lacking time to rearrange itself, collapses in a heap. Now, since the bridge would not have collapsed at all if the bolt had snapped less abruptly—we can even assume that this would have resulted in a *stabler* overall comportment—it was not the bolt's snapping per se that caused the collapse. Adequacy explains this as follows: given the unlikelihood of the molecular mishap, if the snapping had not occurred, it might well not have been sudden if it had; hence the bridge might well not have collapsed. Speaking then of how things *would* have been if not for the snapping, it *cannot* be said that if it had occurred, so would have the bridge's collapse.<sup>44</sup> In other words, the snapping per se was not *adequate* for the collapse; and that is why it was not the cause.

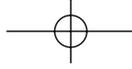
Again the underlying mechanism is worth noticing, both for its own sake and for the connection it suggests between  $x$ 's essence and its causal powers. Properties

<sup>43</sup> Unlike (A<sub>1</sub>) and (A<sub>2</sub>), (A) is not formally dual to (C). But it *is* in an obvious sense dual to:

(C\*) if  $x$  had occurred, then if it had not occurred,  $y$  would not have occurred either.

And since (C\*) is equivalent to (C) in worlds where  $x$  exists, they are interchangeable as conditions on causation.

<sup>44</sup> See Note 38<sup>7</sup> for the relation between would- and might-conditionals. I emphasise that the deterioration begins only *after* the snapping is under way because I want it to be clear that *that very snapping* could have been less abrupt (as opposed to: a less abrupt snapping could have occurred in its place). To deny this would be to say that the snapping, once begun, *could not* have continued apace, i.e., that the impending acceleration was *essential* to it. As for the further claim that if the snapping had not occurred, it *might* have been less abrupt if it had, suppose if you like that indeterminism holds, and that the mishap's objective probability, conditional on preceding events, was vanishingly small.



accidental to  $x$  are potentially ones that it lacks in  $u$ , the  $x$ -including world most similar to the  $x$ -excluding world  $v$  most similar to actuality. In proportion then as  $x$ 's causally important properties are accidental to it, the chances increase of its lacking, in  $u$ , some of the properties by which  $y$  was caused. This raises in turn the likelihood that  $y$  is absent from  $u$ , i.e., that  $x$  is inadequate for  $y$ . So, adequate causes cannot have too many of their causally important properties only accidentally.

## 9. EFFECTS AS REQUIRING THEIR CAUSES

When a cause is essentially overestimated, as, e.g., when Xanthippe blames Socrates's death on his guzzling the hemlock, this often shows itself in a violation of the contingency condition. Not always, though. Imagine that Socrates is a sloppy eater who infallibly guzzles what he drinks. Then his death *might*, I suppose, be contingent on his guzzling the hemlock; but Xanthippe is as unconvincing as ever when she calls it the effect of his doing so. Or imagine that Poindexter, his testimony complete, attends the symphony, where his talking so irritates his fellow concert-goers that he is ejected from the hall; and moreover that, although this plays no role in his ejection, he knows what he is saying to be untrue. To attribute Poindexter's ejection to his *lying* as opposed to his *talking* seems hardly credible. Yet if Poindexter is talking only to pass along misinformation, the ejection may well be contingent on both.

Where do these attributions go wrong? In both cases one wants to say that not *all* of the proposed cause was needed. Included in the guzzling, for example, was a lesser event, the drinking, which would still have done the job even in the guzzling's absence. By hypothesis, of course, without the guzzling this lesser event would not have occurred; but that doesn't stop us from asking what would have happened if it had, and evaluating the guzzling on that basis. Accordingly we define  $x$  as *required* for  $y$  iff:

- (R) Given any  $x^-$  strictly weaker than  $x$ , if  $x^-$  had occurred without  $x$ ,  $y$  would not have occurred.

Among its other advantages (see below), (R) gives the intended result that Socrates's guzzling the hemlock is not what killed him. For the guzzling was required for the death only if there was no strictly weaker event such that, if it had occurred in the guzzling's absence, the death would still have ensued. Socrates's drinking the hemlock being a counterexample to this, his guzzling the hemlock is rejected as not required.

Against the essentiality of causally irrelevant properties, I complained that too much of it jeopardises the contingency of effect on cause. But the argument had a loophole. All that follows from  $x$ 's possessing causally irrelevant properties essentially is that there are *some* worlds from which  $x$  is absent for causally





irrelevant reasons. This leaves it open that in the *nearest*  $x$ -less world  $v$ ,  $x$ 's absence is for failure of one or more of its causally *relevant* properties. In that case, we would expect  $y$  not to occur in  $v$ . So the threatened conflict with contingency need not always materialise.

With condition (R), this loophole can be partly closed. Remember that (C) concerns itself with the nearest world from which  $x$  is absent *for whatever reason*. Subject though to the availability of suitable weakenings,<sup>45</sup> (R) shifts the focus to the nearest worlds from which  $x$  is absent *specifically* for lack of causally irrelevant properties. For  $y$  to be missing from the former world is understandable, but its existence should *not* be threatened if, as in the latter worlds, properties are lacking on which it is not in any case causally dependent.<sup>46</sup> So essential-but-irrelevant properties are likelier to result in violations of (R) than of (C); this consolidates our earlier conclusion that properties of the cause unneeded by its effect cannot be too often essential to it.

## 10. CAUSES AS ENOUGH FOR THEIR EFFECTS

Adequacy was used to explain why the bolt's snapping *per se* could not be blamed for the bridge's collapse. But this required a special assumption: that if the snapping had not occurred, it might well not have been sudden if it had. Suppose instead that when the temperature is extremely low, as on this occasion, soft bolts snap suddenly if at all. Barring an implausible counterfactual dependence of the temperature on the bolt, it would have been just as cold if the snapping had not occurred. But then, given the effect of cold on soft bolts, if the snapping *had* occurred, it would still have been sudden, and the collapse would still have followed. Since now the snapping *is* adequate for the effect, the problem with taking it for the cause lies elsewhere; and the obvious thought is that although the snapping was *part* of the cause, the effect required more. Suppose we call  $x$  *enough* for  $y$  if:

- (E) For all (actually occurring)  $x^+$  strictly stronger than  $x$ ,  $y$  does not require  $x^+$ .<sup>47</sup>

<sup>45</sup> See Sections 11–12.

<sup>46</sup> Here is the argument more explicitly: let  $x^-$  be the result of deleting some set  $I$  of causally irrelevant properties from  $x$ 's essence, and consider what happens in the nearest world  $w$  in which  $x^-$  occurs in the absence of  $x$ . By (1)–(3) of Section 4,  $x$  is absent from a world iff  $x^-$  is either nonexistent, or lacks some  $I$ -property, there. Since  $x^-$  *does* occur in  $w$ ,  $w$  is the nearest world in which  $x^-$  occurs without some  $I$ -property. But then the question whether  $y$  occurs in  $w$  is the question whether it would have occurred, if  $x^-$  had been without some  $I$ -property; and since the  $I$ -properties are by hypothesis irrelevant to  $y$ 's production, the answer must presumably be that it would. Thus there is an event weaker than  $x$  such that  $y$  would still have occurred if that event had occurred in  $x$ 's absence; it follows that  $x$  is not required by  $y$ .

<sup>47</sup> This notion of enoughness is prefigured in Dretske and Snyder (1973) and Anscombe (1975).



Because the bolt's *suddenly* snapping was required for the bridge's collapse, its snapping per se was not enough.

Adequate causes, I said, cannot have too many of their causally relevant properties accidentally. However, the argument I gave was not airtight. There is a conflict with adequacy only if  $x$  lacks some causally relevant property in world  $u$  specifically (as before,  $u$  is the nearest  $x$ -including world to the nearest  $x$ -omitting world to actuality). But all that follows from such a property's being accidental to  $x$  is that  $x$  lacks it *in some world or other*; and why should the property choose world  $u$  to put in its nonappearance? Fortunately what happens in  $u$  is not decisive, where the enoughness condition is concerned: subject to the availability of suitable strengthenings and weakenings,<sup>48</sup> enoughness homes in on the nearest worlds in which  $x$ 's relevant but accidental properties *do* fail. These being worlds in which the effect is unlikely to eventuate, this increases the pressure on causally relevant properties not to be accidental.<sup>49</sup>

## 11. PROPORTIONALITY AND CAUSAL ONTOLOGY

By *proportional* events, I mean events satisfying the contingency, adequacy, requirement, and enoughness conditions:

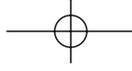
- (C) (If  $x$  had occurred, then) if  $x$  had not occurred, neither would have  $y$ ;
- (A) If  $x$  had not occurred, then if  $x$  had occurred,  $y$  would have occurred as well;
- (R) For all  $x^-$  strictly weaker than  $x$ , if  $x^-$  had occurred without  $x$ ,  $y$  would not have occurred; and
- (E) For all (occurring)  $x^+$  strictly stronger than  $x$ ,  $x^+$  is not required for  $y$ .

(Call the conjunction of these conditions (P).) Whether  $x$  is proportional to  $y$  is sensitive, I have been arguing, to the content of  $x$ 's essence, specifically to how well its essence lines up with the properties by which  $y$  was brought into being. To the extent then that causes are proportional to their effects,  $x$ 's essence bears similarly on its causal powers.

<sup>48</sup> See Sections 11–12.

<sup>49</sup> Here is the argument in full: assume towards a contradiction that  $x$  is required by, and enough for,  $y$ , although  $x$  has causally relevant properties  $R$  only accidentally; and let  $x^+$  come from  $x$  by expanding the latter's essence to include these  $R$ -properties. Then a case can be made that  $y$  requires  $x^+$ , too. An event  $(x^+)^-$  strictly weaker than  $x^+$  is an event whose essence falls short of  $x^+$ 's by some combination  $S$  of  $R$ -properties and properties in  $x$ 's essence. Consider the nearest world  $w$  in which  $(x^+)^-$  occurs without  $x^+$ . By (1)–(3) of Section 4,  $(x^+)^-$  lacks some or all of the  $S$ -properties in  $w$ . But the  $S$ -properties are predominantly causally relevant to  $y$  (the  $R$ -properties by hypothesis, and the properties from  $x$ 's essence because  $y$  requires  $x$ ). Probably then  $y$  does not occur in  $w$ . Thus for an arbitrary weakening  $(x^+)^-$  of  $x^+$ ,  $y$  would probably not have occurred if  $(x^+)^-$  had occurred without  $x^+$ . Assuming (!) that each of these probable counterfactuals is in fact true,  $y$  requires  $x^+$ . But this is contrary to our assumption that  $x$  was enough for  $y$ .





At this point it seems that we have answered the sceptical challenge of Section 3: what possible difference can an event's essence make to what it causes? Yet a major issue has been dodged. Up to now I have been talking as though proportionality was a single well-defined condition. But from (P)'s logical form one sees that its demands *intensify* as its quantifiers range over more and more events. Equivalently we could say that the condition (P) expresses is a monotonically increasing function of its quantificational domain—what I will call *causal ontology*. The question is, how strong or weak a condition can (P) be made to signify by variation in this ontology?

Imagine an ontology so meagre that events have no coincidents but themselves ( $x_1$  is coincident with  $x_2$  only if they are identical). Since strengthening entails coincidence, none of these events is strictly stronger than any other. (R) and (E) are therefore trivialised, and (P) collapses into (C) and (A), the contingency and adequacy conditions. So that is one extreme. For the other, suppose we call  $x$  *necessary* for  $y$  iff  $y$  cannot—metaphysically cannot—exist without it, and *sufficient* for  $y$  iff  $x$  cannot exist without  $y$ . Unless  $x$  is necessary and sufficient for  $y$ , it turns out, there is room, reliably exploitable by an ontologically unscrupulous monkey wrencher, for a *counterexample* to the hypothesis that  $x$  is required by and enough for  $y$ .<sup>50</sup> Depending on causal ontology, then, (P) can mean as little as contingency-plus-adequacy, or as much as necessity-plus-sufficiency.

Now, I take it that neither of these extremes is tolerable: the second makes (P) absurdly overdemanding,<sup>51</sup> and on the first the commensuration ideal is all but abandoned. How, though, to find the happy medium? Suppose we conceive the problem operationally: starting from a modest foundation, say an ontology with no nontrivial coincidence relations, and building upwards, does there emerge at some point a natural brake on the construction? The more ontology grows, we know, the more commensurate (P)-related events become.

<sup>50</sup> Let  $x$  be unnecessary for  $y$ . Then the set  $W$  of worlds in which  $y$  occurs but  $x$  does not is nonempty. Supposing a suitably rich ontology,  $x$  can be weakened to an event  $x^-$  existing in all the  $x$ -worlds plus  $W$ . Since  $W$  contains every world in which  $x^-$  occurs without  $x$ , if  $x^-$  had occurred in  $x$ 's absence, that would have been in some  $W$ -world. But  $y$  occurs in *every*  $W$ -world, so  $y$  would still have occurred if  $x^-$  had occurred without  $x$ . Thus  $x^-$  is a counterexample to the hypothesis that  $y$  requires  $x$ . That completes the argument that (R) entails necessity on the condition of an unrestricted ontology. Next we argue that (R) and (E) entail sufficiency on the same condition. Assume for contradiction that although  $x$  and  $y$  satisfy (R) and (E),  $x$  is insufficient for  $y$ . By the previous result, we can assume that  $x$  is necessary for  $y$ . Because  $x$  is not sufficient for  $y$ , the set  $W$  of worlds in which  $x$  occurs but  $y$  does not is nonempty. Ontology being unrestricted,  $x$  has a strengthening  $x^+$  which differs from  $x$  only in being absent from these  $W$ -worlds, i.e.,  $x^+$  occurs in exactly the  $x$ -worlds which contain  $y$ . Since  $x$  is necessary for  $y$ , every  $y$ -world is an  $x$ -world; and since  $x^+$  exists in every world containing both  $y$  and  $x$ , every  $y$ -world is an  $x^+$ -world as well. From this last it follows that  $y$  requires  $x^+$ . But then  $x$  is not enough for  $y$ , contrary to assumption.

<sup>51</sup> Even if an  $x$  necessary and sufficient for  $y$  could be found, they would not be "distinct existences" in the sense of each being possible without the other. On most theories this rules out a causal relation between them (cf. Hume 1968, Bk. I, Part III, Secs. III and XII; and Mackie 1974, Ch. 1).



What we'll discover is that too *much* commensuration is a bad thing; and this for reasons that make themselves felt *before* we arrive at the upper extreme just noted.

To fix ideas, consider the course our process takes with a specific causal episode, say, Lucy's demolishing a sandcastle with a rock. Relative to the going ontology, we find, say, that  $x = \textit{her dropping that rock}$  is proportional to  $y$ , the castle's breakup. Now let the pool of events be gradually enlarged. As (P)'s demands increase, we are forced to nominate new causes in place of the old, each momentarily satisfying us until further events make their way onto the scene and the cycle repeats. For instance, the sequence  $\langle x_i \rangle$  of causes might begin like this (the letters in the right-hand column indicate whether (R) or (E) motivated  $x_i$ 's rejection);

$x_1 = x = \textit{her dropping that rock}$	(R)
$x_2 = \textit{her dropping a rock}$	(E)
$x_3 = \textit{her dropping a large rock}$	(R)
$x_4 = \textit{her dropping a large object}$	(E)
$x_5 = \textit{her dropping a large object from above the sandcastle}$	(R)
$x_6 = \textit{her propelling a large object in the direction of the sandcastle}$	(E)
$x_7 = \textit{her propelling a large and heavy object in the direction of the sandcastle}$	(R)
$x_8 = \textit{the propulsion of a large and heavy object in the direction of the sandcastle}$	(E)
$x_9 = \textit{the propulsion of a large, heavy, stable object at a good velocity in the direction of the sandcastle}$	(R)

Of course, even  $x_9$  will not satisfy us for long. How slowly the object can afford to be moving depends on how heavy and stable it is; and how heavy it needs to be depends likewise on its size and velocity, whose permissible values depend in turn on the gravitational and other forces then in effect. So even at this relatively early stage we seem driven to something like:

$x^* =$  The propulsion of a suitably large and heavy, sufficiently stable object at an adequate velocity towards the sandcastle in the presence of appropriate gravitational forces and the absence of effective electromagnetic interference,

where 'suitable', 'sufficient', 'adequate', 'appropriate', and so on, are to indicate that these parameters should assume mutually satisfactory values relative to the goal of achieving the castle's collapse. Describing events like this as *dedicated* to their effects helps to underline that their conditions of occurrence are of the

form (exaggerating somewhat): there obtains *some* such combination of the given factors as will result in the effect's occurrence. What we are after is a rationale for excluding dedicated events from our ontology.

Associated with Aristotle is the idea that certain outcomes, what we might ordinarily call 'accidents', are not as such caused. To borrow one of his examples, there may be a cause for your entering the market at 4 p.m., and a cause for your debtor's entering it then; we may even suppose that these combine to form a cause of your *both* entering it at 4 p.m. But if your meeting was, as we say, *accidental*, we would precisely *not* expect to find a cause for your entering the market *at the same time*.<sup>52</sup> Neither do we expect a cause if, as in the O. Henry story, someone receives as a Christmas present the very thing that she can no longer use. There may be causes for her receiving what she did, and for her changed situation, but nothing accounts, on an intuitive level, for their discord *as such*.

Although I have no foolproof definition to offer, it seems characteristic of accidents that they essentially specify the *relations* among several causally independent parameters but without prejudice to their separate values (e.g., that you enter at the same time, but not the time you enter). Rough as it is, this suggests that our reluctance to ascribe causes to accidents is what one would anyway expect from the commensuration principle. For what would an accident's commensurate cause be like? On the one hand, it would have to arrange for each of the causally disconnected factors in whose rapport the effect consists to respect the others; on the other hand, it could not fix these separately on pain of overshooting the mark. The problem is to see how the first condition can be met without sacrificing the second.<sup>53</sup>

Accidents do not, *per se*, have causes.<sup>54</sup> But accidents are what we are dealing with, in causes dedicated to their effects. Take, for instance, the strange event  $x^*$ : the propulsion of a suitably large and heavy, sufficiently stable object at an adequate velocity towards the sandcastle in the presence of appropriate gravitational forces and in the absence of effective electromagnetic interference. As with the other accidents mentioned,  $x^*$  occurs iff there obtains *some* such combination of factors as meets a certain externally imposed condition (in this case, that of securing a certain effect). What we said about the other accidents therefore applies here: it is obscure how any prior event could hope to coordinate these factors without constraining them beyond what  $x^*$  requires.

<sup>52</sup> Here I assume that your entering the market at the same time is, to the extent we can make sense of it as a token event at all, something that *could* have occurred (e.g.) at 4:03 p.m.; in this it differs from your both entering the market at 4 p.m.

<sup>53</sup> This is compatible with there being antecedent events which causally *necessitate* the accident; what we are looking for is a *commensurate* antecedent. For discussion, see Kim (1974), Sorabji (1980), and Lewis (1986a, esp. Sec. VII).

<sup>54</sup> Or, if a cause was for some reason insisted on, we would expect it to be of an even more *outré* variety than the accident itself, and so heir to the same difficulties in more aggravated form.

Along with the problem of finding causes for dedicated events, there is a problem finding effects for them to cause. Thus our original cause  $x$  (Lucy's dropping the rock) is at least *roughly* commensurate with not only  $y$  (the collapse) but any number of other events: the sensation of release, the twins' cry of alarm, the honeybee's sudden flight, etc.; and it goes with this that relative to a moderate, although not an extravagant, causal ontology,  $x$  will come out proportional to all of them. Not so with  $x^*$ , which stands little chance of proportionality with any but the collapse. It is not required, because, e.g., it makes no difference to the honeybee's flight if the projectile is so unstable that it disintegrates just on reaching the castle; and it is not enough, because, e.g., there would have been no sensation, if the propulsion had been accomplished by mechanical means. So the contemplated additions to causal ontology reduce overall effectiveness in *two* ways: both by undermining preexisting causal relations (like that between  $x$  and  $y$ ); and by their own relative ineffectiveness ( $x^*$  causes little else *but*  $y$ ).

Summing up, as events are multiplied in the interests of causal precision, they suffer in accountability on the side of their causes, and versatility on the side of their effects. As a whole the causal order becomes fragmentary and disconnected; ultimately we find ourselves in a world whose every outcome derives from an unmoved mover dedicated precisely to it. Yet with too few events proportionality cannot carry out its assigned task of enforcing commensuration. I conclude that the right ontology, for purposes of causal theory, is the one that strikes the best overall compromise between commensuration on the one hand, and the unity and integrity of the causal order on the other.

## 12. WORLD-DRIVEN VS. EFFECT-DRIVEN CAUSES

“Surely, though, an objectively *ideal* compromise is not to be hoped for; so the above leaves the question of causal ontology partly open”. So much the better, I argue now.

Inspiring the commensuration constraint is a certain platitude: the cause is the thing that ‘made the difference’, in the obtaining circumstances, between the effect's occurring and its not. But the platitude can mean more than one thing, according to which of two related contrasts we want the cause to mark. First is the contrast between worlds where the effect goes on to occur and those where it doesn't:  $x$  is to be the choice-point, as it were, between these two types of future. Second is that between the *actual* world and worlds where the effect does not occur:  $x$  is to indicate how the choice of a  $y$ -containing future is implemented *here* as opposed to elsewhere.<sup>55</sup>

<sup>55</sup> In this and the next few paragraphs, ‘worlds’ means: worlds agreeing with the actual world in contextually determined background conditions (what the platitude called the obtaining circumstances). The notion admittedly bears very little scrutiny and I wish I knew how to express my point without it.

Mill gives the example of a man dying from a bad meal: “[I]f a person eats of a particular dish, and dies in consequence, that is, would not have died if he had not eaten of it, people would be apt to say that his eating of that dish was the cause of his death”.<sup>56</sup> But is the cause his eating *poisonously tainted* oysters, or his eating *those* oysters? Apart from context, it could be either. Which way we go depends on which of the just-mentioned contrasts we have mainly in mind. Intent of the first contrast, and concerned to find the antecedent that marks off the effect-worlds from the others, we look for an *x* which essentially involves what it *took* for the effect to occur: in this case, the man’s eating poisonously tainted oysters, never mind exactly which. Like any event, *x* occurs in some determinate way, but its essence homes in on those aspects of its occurrence critically implicated in *y*’s production. But suppose our aim is to say what *specifically* happened in the *actual* world, to make it one of the worlds in which *y* occurred. Then we look for an event which brings out how actuality contrived to *realise* these critical aspects: in this case, the man’s eating *those* oysters. So, where the first sort of cause emphasises what the effect *needed* in order to occur, the second indicates something of *how its needs were in fact met*.<sup>57</sup>

Two elements have been distinguished in causal judgement. Both are present, to greater or lesser degree, practically whenever we nominate one event as another’s cause. Where the first element predominates, and *x* is tailored to the effect’s causal requirements, I call the judgement *effect-driven*; where the second element predominates, and *x* is considerate of how those requirements are in fact fulfilled, I call it *world-driven*. So to blame the man’s death on his eating *tainted* oysters is to make an effect-driven judgement; the retort that it was his eating *those* oysters that killed him is world-driven. Again, your judgement is effect-driven, if you attribute Rumpelstiltskin’s furious stamping to the miller girl’s guessing his name, or Icarus’s fall to his flying so near the sun; world-driven, if you propose instead her saying “Rumpelstiltskin”, or his flying so high.<sup>58</sup>

Assuming that neither of these attributional styles is to be privileged, how on the present theory can we make room for both? This is where causal ontology returns to do useful work. Enlarging it, we saw, turns up the commensuration

<sup>56</sup> Mill (1950, Bk. III, Ch. V, §3).

<sup>57</sup> Related to this, causes of the first sort will be more *robust* than those of the second, in the sense of continuing to operate through a broader range of counterfactual cases. Compare Putnam’s notion of an “autonomous explanation” in ‘Philosophy and Our Mental Life’: “The same explanation will go in any world (whatever the microstructure) in which [the same] *higher level structural features* are present. In that sense [the] *explanation is autonomous*” (Putnam 1975b, p. 296).

<sup>58</sup> Two remarks. First, the world-driven/effect-driven distinction is a relative one; some causes are more world-driven than others, but none is world-driven in an absolute sense. Second, as the examples show, there is no direct correlation between a cause’s world-drivenness and its strength. What does happen as causes become more world-driven is that their essences become more explicit about how the effect’s needs were in fact met. But this often brings with it a *loss* of information about what those needs were, and so about how it was that what actually happened served to meet them. So although there is strengthening along one dimension there may well be weakening along another.

pressure on would-be causes. Relatively incidental features of the causal scene, distinctive though they might be of the actual progress of events, are worn away to reveal the steadier causal currents beneath. Such a strategy can of course be taken too far (a theme of the last section). Practised in moderation, though, it brings an agreeable broadening and deepening of causal judgement, what I described by saying that these judgements become less world- and more effect-driven. Sometimes, it is true, we are willing to accept a shallower causal story in return for more discriminating information about what took place; in that case an easing of commensuration pressures is called for and hence a reduced causal ontology. So, if the question is: isn't ontology something to be settled uniquely, and identically across applications?—my reply is that this is a common assumption but not always a useful one. Underlying as it does a familiar and advantageous flexibility of causal judgement, the openness of *causal* ontology, at least, is all to the good.

### 13. EPIPHENOMENALISM<sup>59</sup>

Writing to Descartes in 1643, Princess Elisabeth requested an explanation of “how man’s soul, being only a thinking substance, can determine animal spirits so as to cause voluntary actions”.<sup>60</sup> Dualism has been struggling to dissociate itself from epiphenomenalism ever since. The outlines of the problem are clear enough: if mind and body are metaphysically separate, as the dualist says, then how can the one affect the other? Three centuries of dualist apologetics on the topic have failed to provide an answer.

Why though should this old problem concern anyone today? Dualism is an evolving doctrine, and its Cartesian version has by now given way to something far less outlandish, to which Elisabeth’s original complaint about the obscurity of cross-category interaction no longer applies. Immaterial minds are gone, and although mental *phenomena* (facts, properties, events) remain, the contemporary dualist admits, in fact insists, that they are physically realised. All that survives from Cartesianism is the denial of their numerical *identity* with their physical bases. Surely it would be hard to imagine a dualism more congenial to mind/body causation than this!<sup>61</sup>

Indeed it would. But epiphenomenalism has been evolving too, and in its latest and boldest manifestation, this is all the dualism it asks for. The paradoxical

<sup>59</sup> Parts of this and the next two sections are based on Yablo (1992).

<sup>60</sup> Wilson (1969, p. 373).

<sup>61</sup> In case it seems odd to describe the theory just sketched as dualistic, I should explain that all I mean by the term is that mental and physical phenomena are, contrary to the identity theory, *distinct*, and contrary to eliminativism, *existents*. That this much dualism is acceptable even to many materialists is in a way the point. Having broken with Cartesianism over its troubles with mind/body causation, they find to their horror that epiphenomenalism lives equally happily on the lesser dualism latent in their own view.

result is that, at a time when the prospects for accommodating mental causation seem little short of ideal, epiphenomenalist anxiety runs higher than ever. Nor is this a pretended anxiety, put on for dialectical purposes but posing no genuine danger to established views. Some say we must simply make our peace with the fact that “the mental does not enjoy its own independent causal powers”.<sup>62</sup> Others would renounce (distinctively) mental entities altogether, rather than see them causally disabled.<sup>63</sup>

Radical as these proposals are, they are backed by a rather straightforward line of thought: “How can mental phenomena make any causal difference to what happens physically? Every physical outcome is causally assured by its physical antecedents; its mental antecedents are therefore left with nothing further to contribute”. This is the *exclusion* argument for epiphenomenalism. Here is the argument as it applies to mental events; for the version that applies to properties, replace ‘event  $x$ ’ with ‘property  $X$ ’:<sup>64</sup>

- (1) If an event  $x$  is causally sufficient for an event  $y$ , then no event  $x^*$  distinct from  $x$  is causally relevant to  $y$  (*exclusion*).<sup>65</sup>
- (2) For every physical event  $y$ , some physical event  $x$  is causally sufficient for  $y$  (*determinism*).<sup>66</sup>
- (3) For every physical event  $x$  and mental event  $x^*$ ,  $x$  is distinct from  $x^*$  (*dualism*).
- (4) So: for every physical event  $y$ , no mental event  $x^*$  is causally relevant to  $y$  (*epiphenomenalism*).

This is bad enough—as Malcolm says in ‘The Conceivability of Mechanism’, it calls into question even the possibility of speech and action—but a simple extension of the argument seems to deprive mental phenomena of all causal influence

<sup>62</sup> Kim (1983, p. 54).      <sup>63</sup> Schiffer (1989, ch. 6).

<sup>64</sup> So ‘ $x$ ’ and ‘ $x^*$ ’ become ‘ $X$ ’ and ‘ $X^*$ ’, and where either is prefixed by ‘event’, this becomes ‘property’; ‘event  $y$ ’ and ‘event  $z$ ’ are unaffected. Although causes and effects are events, properties as well as events can be causally relevant and/or sufficient. I try to remain neutral about what exactly causal relevance and sufficiency come to, e.g., causal sufficiency could be absolute, or it could be sufficiency-in-the-circumstances. Versions of the exclusion argument are found in Malcolm (1968/1982), Goldman (1969), Campbell (1970), Honderich (1982), and Kim (1979, 1989). Analogous objections are sometimes raised against the causal claims of other phenomena apparently unneeded in fundamental physical explanation, e.g., macro- and colour-phenomena. The next few sections offer a potentially general strategy of response.

<sup>65</sup> Some authors use a slightly different premise: if  $x$  is causally sufficient for  $y$ , then *barring overdetermination*, no  $x^* \neq x$  is causally relevant to  $y$ . I do not consider this form of the argument explicitly, but my response will be easy to guess from what I say about the version in the text.

<sup>66</sup> Although (2) could obviously be questioned, I take it that physical determinism isn’t the issue. For one thing, the conviction that mind makes a causal difference is not beholden to the contemporary opinion that determinism is false, and would remain if that opinion were reversed. Second, nothing essential is lost if ‘ $x$  is causally sufficient for  $y$ ’ is replaced throughout by ‘ $x$  fixes  $y$ ’s objective probability’. So unless the argument can be faulted on other grounds, mental causation is problematic under indeterminism, too.

whatsoever. Every event  $z$  of whatever type is metaphysically necessitated by some underlying physical event  $y$ , whose causally sufficient physical antecedents are presumably sufficient for  $z$  as well. But then by the exclusion principle, mental phenomena are entirely causally inert. And now it is not only speech and action that are endangered but also thinking.

Now, it is important that the exclusion argument raises *two* problems for mental causation, one about mental particulars (events) and the other about mental properties. Their evident similarity notwithstanding, philosophers have tended to treat these problems in isolation and to favour different strategies of solution.<sup>67</sup> Easily the most common reaction to the first is to insist that mental events are *identical* with (some among) physical events, whose causal powers they therefore share.<sup>68</sup> Such a response is at best incomplete, because of the second problem. Mental events are effective, maybe, but not in virtue of their mental properties; any causal role which the latter might have hoped to play is occupied already by their physical rivals.<sup>69</sup> Although someone *could*, following the line above, attempt to identify mental with (some among) physical properties, this response is now discredited; the argument bears examination, since, appropriately modified, it seems also to cast doubt on *token* identity.

When philosophers abandoned the hope of finding for every mental property an identical physical property, their reason was that mental properties seem intuitively to be *multiply realisable* in the physical.<sup>70</sup> But some care must be taken about what this means. Sometimes it is claimed that for *any* pair of properties, one mental and the other physical, something could have the first without the second. Really, though, this is stronger than intended or needed. Imagine a philosopher who holds that necessarily every thinker is spatially extended. Presumably she could also accept multiple realisation, intuitively understood, without falling into inconsistency. But, since the necessitation of extension by thinking is the necessitation of a physical property by a mental one, her view actually runs contrary to multiple realisation as just explained! *Provided that they are suitably unspecific*, then, physical properties *can*, compatibly with multiple realisation, be necessitated by mental properties; which suggests as the thesis's

<sup>67</sup> An exception is Kim (1984b).

<sup>68</sup> In his (1970/1980a), Davidson advances the token-identity theory as the solution to a different problem: singular causal claims need always to be backed by strict causal laws; strict laws are always physical laws; physical laws subsume physical events only; therefore mental events are inefficacious, unless they are also physical events.

<sup>69</sup> Again, this needs to be distinguished from a quite different worry directed mainly at Davidson's (1970/1980a) anomalous monism: singular causal claims always need to be backed by some strict causal law;  $x$ 's causally relevant properties vis-à-vis  $y$  are those figuring in the antecedent of some such backing law; strict laws never involve mental properties; so  $x$ 's mental properties are causally irrelevant. For discussion, see Stoutland (1980), Honderich (1982), Loewer and LePore (1987, 1989), Fodor (1989), Macdonald and Macdonald (1986), and McLaughlin (1989) (some of these papers discuss the exclusion objection also). Note that the exclusion objection assumes nothing about the role of laws in causation or in the characterisation of causally relevant properties.

<sup>70</sup> See, e.g., Putnam (1980) and Block and Fodor (1972/1980).

proper formulation that  $M$  necessitates no physical properties specific enough to necessitate  $M$  in return:

- (M) Necessarily, for every mental property  $M$ , and every physical property  $P$  that necessitates  $M$ ,  $P$  necessitates  $M$  asymmetrically, i.e., possibly something possesses  $M$  but lacks  $P$ .

For purposes of refuting the type identity theory, note, (M) is all that's needed. Assume for contradiction that  $M$  is  $P$ . Then  $P$  necessitates  $M$ . But then by (M), a thing can have  $M$  otherwise than by way of possessing  $P$ , contrary to their assumed identity.

What is not often noticed is how easily the above adapts to mental and physical events. Take, for instance, a pain sensation  $s$ , and some underlying brain event  $b$  alleged to be identical with  $s$ ; and grant the identity theorist that  $b$  at least strengthens the pain. The problem is that as  $b$  takes on the degree of essential physical detail without which the pain is not necessitated, the likelihood increases that the pain is possible even in  $b$ 's absence. Something like this is one of Kripke's arguments against token identity:

[*B*]eing a brain state is evidently an essential property of  $b$  (the brain state). Indeed, even more is true: not only being a brain state, but even being a brain state of a specific type is essential to  $b$ . The configuration of brain cells whose presence at a given time constitutes the presence of  $b$  at that time is essential to  $b$ , and in its absence  $b$  would not have existed. Thus someone who wishes to claim that the brain state and the pain are identical must argue that the pain could not have existed without a quite specific type of configuration of molecules.<sup>71</sup>

*Prima facie*, it seems obvious that the pain could still have occurred, even if that specific configuration of molecules hadn't; and, as Kripke says, the *prima facie* appearances aren't easily defeated. But if the molecular configuration is essential to  $b$  alone, then  $b$  strengthens  $s$  *properly* or *asymmetrically*. Extended across mental and physical events in general, this amounts to an analogue for particulars of the multiple realisability thesis:

- (m) For every mental event  $m$ , and every physical event  $p$  which strengthens  $m$ ,  $p$  strengthens  $m$  asymmetrically.

By (m), any physical  $p$  specific enough to *strengthen* a mental event  $m$  is *too* specific to be *identical* with  $m$ . Token dualism follows, by the same reasoning as before.

Isn't this playing into the epiphenomenalist's hands, though? If  $m$  is distinct from  $p$ , then  $m$  can influence an outcome only to the extent that  $p$  leaves that outcome causally undecided. Effects which  $p$  causally guarantees, then, it renders insusceptible to causal influence from any other source,  $m$  included. Assuming,

<sup>71</sup> Kripke (1980, p. 147–48) with inessential relettering.

for instance, that all it took for me to wince, clutch my brow, and so on, was my antecedent physical condition, everything else was strictly by the way. Since my headache is a different thing from its physical basis, it is not a *bona fide* causal factor in my headache behaviour.

How plausible we find this argument depends on how much rivalry we admit between an effect's would-be causal antecedents. Does  $x$ 's causal sufficiency for  $y$  really make *all* of  $y$ 's other antecedents irrelevant? Such a view implies, absurdly, that  $y$  owes nothing to  $x$ 's causal antecedents, or to the causal intermediaries by which  $x$  generates  $y$ .<sup>72</sup> At least as it applies to events, then, the exclusion principle is overdrawn; but not, or not yet, in a way that helps with the problem of mental causation, for the charge against mental causes is that they are preempted by underlying physical causes to which they are bound, not causally, but in some more intimate metaphysical association. Next we consider what their relation could be, that events related in *that* way do not compete for causal influence.

For the reasons given, I find no fault with type- or token-dualism, or with the picture of mental phenomena as necessitated by physical phenomena which they are possible without. Rather than objecting, in fact, to the asymmetric necessitation picture, I propose to go it one better. It will be easiest to begin with mental and physical properties. According to a still reputable traditional doctrine, some properties stand to others as *determinate* to *determinable*, e.g., *scarlet* is a determinate of *red*, *red* is a determinate of *coloured*, and so on. Since the distinction is relative, one does better to speak of a determination *relation*, where:

( $\Delta$ )  $P$  determines  $Q$  iff to be  $P$  is to be  $Q$ , not *simpliciter*, but in a specific way.

As traditionally understood,<sup>73</sup> determination involves conceptual and metaphysical elements jumbled confusingly together. Metaphysically, the main idea is that:

( $\Delta$ )  $P > Q$ <sup>74</sup> only if: (i) necessarily, for all  $x$ , if  $x$  has  $P$  then  $x$  has  $Q$ ;  
(ii) possibly, for some  $x$ ,  $x$  has  $Q$  but lacks  $P$ .

Not always distinguished from this is a requirement of asymmetric *conceptual* entailment: there is no conceptual difficulty about  $Q$ s which are not  $P$ s, but the reverse hypothesis is conceptually incoherent.

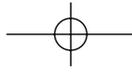
Now, just as the discovery by Kripke and Putnam of *a posteriori* necessities upset the conceptual equivalence condition on property-identity,<sup>75</sup> it also invites

<sup>72</sup> Goldman (1969) and Kim (1989) make related observations.

<sup>73</sup> Johnson (1964, ch. 11), and Prior (1949) are classic discussions.

<sup>74</sup> ' $P > Q$ ' is short for:  $P$  determines  $Q$ .

<sup>75</sup> For example, the property of being salt is identical to the property of being sodium chloride, but it is not conceptually necessary that all and only salt is sodium chloride. See Putnam (1975a, p. 306); Kripke (1980, pp. 115ff.); and Yablo (1992).



a reconsideration of the conceptual condition on *determination*. Let  $K$  be some highly specific micromechanical property chosen so that necessarily, whatever is  $K$  is at temperature  $90^\circ\text{C}$ . Assuming that warmer or cooler  $K$ s cannot be ruled out on conceptual grounds alone,  $K$  does not determine the temperature property, at least not in the full-blown traditional sense. Or let the pertinent aspects of my physical condition be encoded in some physical property  $P$ , such that unthinking  $P$ s are metaphysically impossible. Again, barring some unsuspected conceptual entailment from physics to thought, traditional determination fails.

Yet the relevance of these *conceptual* possibilities to the properties' *metaphysical* relations is obscure; and since it is only the metaphysics that matters to causation, it seems wisest simply to drop the second, epistemological, component of the traditional doctrine, and to conceive determination in purely metaphysical terms.<sup>76</sup> This opens the way to treating  $K$  as a determinate of the temperature property, and (what is of course the point)  $P$  as a determinate of thinking.

Then why not see all mental properties as determinables of their physical bases? Such a view is in fact implicit in the reigning orthodoxy about mind/body relations, namely, multiple realisation (M) plus the supervenience thesis:

- (S) Necessarily, if something has a mental property  $M$ , then it has a physical property  $P$  that necessitates  $M$ .<sup>77</sup>

By (S), anything with a mental property has a necessitating physical property, which by (M), necessitates the mental property asymmetrically. Necessarily, then, something has a mental property iff it has a physical property by which that mental property is asymmetrically necessitated. But this is extremely suggestive, for with 'determines' substituted for 'asymmetrically necessitates' it becomes:

- (D) Necessarily, something has a mental property  $M$  iff it has also a physical determination  $P$  of that mental property;

and (D) is an instance of the standard equation for determinables and determinates generally, viz., that something has a determinable iff it has some determinate falling thereunder. It is hard not to hear this as an argument that, as (D) says, mental/physical relations are a species of determinable/determinate relations.

<sup>76</sup> Thus  $P$  is a determination of  $Q$  just in case the traditional relation's first, metaphysical component is in place, where this consists primarily in the fact that  $P$ s metaphysically must be  $Q$ s, but not conversely. Probably it goes too far to identify determination with asymmetric necessitation outright, otherwise, e.g., conjunctive properties determine their conjuncts and universally impossible properties are all-determining. For dialectical reasons I try to remain as neutral as I can about where determination leaves off and 'mere' asymmetric necessitation begins (Prior (1949) reviews some of the history of this problem).

<sup>77</sup> (S) is Kim's "strong supervenience" thesis (Kim 1984a).



Properties stand in the determination relation iff for a thing to possess the one is for it to possess the other, not *simpliciter*, but in a specific way (this was ( $\Delta$ ) above). But this way of putting things comes naturally, too, in connection with *particulars*, and especially events. If  $p$  is the bolt's *suddenly* snapping, and  $q$  is its snapping *per se*, then for  $p$  to occur is for  $q$  to occur in a specific way, *viz.*, suddenly; likewise for my *slamming* the door to occur is for my *shutting* it to occur, not *simpliciter*, but with a certain forcefulness. Examples like these suggest the possibility of a determination relation for particulars, where:

- ( $\delta$ )  $p$  determines  $q$  iff for  $p$  to exist (in a possible world) is for  $q$  to exist (there), not *simpliciter*, but in a specific way.

As luck would have it, such a relation is already available from Section 4. When one event strictly or asymmetrically *strengthens* another, for the stronger event to occur in a world is for the weaker to occur there, not *simpliciter*, but in possession of the properties by which their essences differ (Section 4, (1)–(3)). So we define  $p$  as a *determination* of  $q$  iff  $p$  strengthens  $q$  asymmetrically.

Perhaps the analogy with properties can be pressed a little further. Corresponding to the multiple realisation thesis (M), we have:

- (m) For every mental event  $m$ , and every physical event  $p$  which strengthens  $m$ ,  $p$  strengthens  $m$  asymmetrically, *i.e.*,  $p$  determines  $m$ .

So far we have no analogue for particulars of (S), the mental/physical supervenience thesis; but suppose, as an experiment, that:

- (s) Whenever a mental event  $m$  occurs, there occurs also a physical event  $p$  that strengthens  $m$ .<sup>78</sup>

There is partial support for this in the supervenience thesis itself. By supervenience, each mental property in  $m$ 's essence is necessitated by some underlying physical property. Even if some or all of these physical properties are only accidental to  $m$ , we can imagine a physical event  $p$  to which they are all essential,

<sup>78</sup> This may seem doubtful, if one insists on seeing  $p$  as (i) a localised brain event, (ii) capable of occurring in isolation from anything like its actual neural context. Imagine a C-fibre stimulation  $b$  realisable in isolated C-fibers afloat in a dish of agar jelly. So realised,  $b$  involves no sensation of any sort, so if  $s$  is a pain sensation, then  $b$  does not necessitate  $s$ , or (therefore) determine it. The moral is not that  $s$  has *no* physical determination, but that (i) and (ii) ask too much. Many mental events seem not to be localisable in any specific portion of the brain. Since determination entails coincidence, their physical determinations are not localisable either (thus  $p$  might be the event of falling into a certain overall neurological condition). Arguably no mental event is localisable, but if  $m$  is an exception, then its physical determination is a localised brain event whose essence is partly extrinsic, *e.g.*, the C-fibers' firing in something like their actual neural environment. (So-called "wide content" mental events raise related problems which I don't discuss. See Fodor (1987, ch. 2; 1991) and Heil and Mele (1991).)



and to which every mental property in  $m$ 's essence is therefore essential, too. Thus every mental property in  $m$ 's essence is also in  $p$ 's essence. Assuming that  $p$  can also be fitted out with essential properties to necessitate what few *nonmental* properties might be found in  $m$ 's essence,  $p$  is the physical strengthening of  $m$  postulated by (s).

Now the analogy is complete. For every mental event  $m$ , (s) guarantees a physical strengthening  $p$ , which by (m) is  $m$ 's determinate. Since the converse is immediate from Section 4, we have:

(d) A mental event  $m$  occurs iff some physical determination  $p$  of  $m$  occurs.

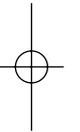
Assuming (m) and (s), the relation of mental to physical events effectively duplicates that of mental to physical properties.

#### 14. DETERMINATION AND CAUSAL RELEVANCE

According to the picture I am promoting, whatever has a mental property  $M$  has also a determining physical property  $P$ , such that to have  $P$  is to have  $M$  in a certain physical way; and whenever a mental event  $m$  occurs, there occurs also a determining physical event  $p$ , such that for  $p$  to occur is for  $m$  to occur in a certain physical way. Yet it is as true as ever that the physical property (event) and its mental counterpart are not the same; and this is all the exclusion objection asked for in the way of mental/physical separation. How then does it respond to the objection to say that the mental item is a determinable of the physical one?

Imagine a pigeon Sophie conditioned to peck at red shapes, and them only; a red triangle is presented, and Sophie pecks. Most people would say that the redness was causally relevant to her pecking, even that this was a paradigm case of causal relevance. But wait! I forgot to mention that the triangle in question was a specific shade of red, say scarlet. Assuming for argument's sake that the scarlet was already causally sufficient for the pecking, the exclusion principle entails that every *other* property was superfluous. So the redness, although it looked to be *precisely* what Sophie was responding to, in reality makes no causal contribution whatever.

Another example concerns properties of events. Suppose that the buildings in a certain region, although built to withstand lesser earthquakes, are in the event of a *violent* earthquake—one registering five or more on the Richter scale—causally guaranteed to fall. When one unexpectedly hits, and the buildings crumble, one property of the earthquake that seems relevant to their doing so is that it was violent. Or so you might think, until I mention that this particular earthquake was *merely* violent, in the sense of registering over five on the Richter scale, but less than six. What with the earthquake's *mere* violence being *already* sufficient for the effect, that it was *violent* cannot have made any causal difference.

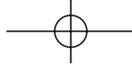


Surprising results, these! To the untrained eye, the redness and the violence are *paradigm cases* of causal relevance, but it takes only a little philosophy to see through them. Yet I take it that our initial reaction was the correct one, and that it is the exclusion principle that is steering us wrong. What the examples really show is, not that the redness and violence are irrelevant, but that determinates do not compete with their determinables for causal influence.

And a good thing, too. For suppose that the competition was real. Then practically whenever a determinable  $Q$  was *prima facie* relevant to an effect, a causally sufficient determinate  $Q'$  of  $Q$  could be found to expose  $Q$  as irrelevant after all.<sup>79</sup> But this would hold equally of  $Q'$ ,  $Q''$ ,  $Q'''$ , etc. So in the end only *ultimate* determinates—properties unamenable to further determination—could hope to retain their causal standing. Or, on second thought, maybe not them either. Not everything about a cause contributes to its effect, and even where a property does contribute, it need not do so in all its aspects. From the examples it is clear that such irrelevancies do creep in, as we pass from determinable to determinate; and if the determination process is continued *ad finem*, they may be expected to accumulate significantly. But then abstracting some or all of this detail away should leave a determinable which, since it falls short of the original only in causally irrelevant respects, is no less sufficient for the effect. By the exclusion principle, this robs even ultimate determinates of their causal powers; and now it begins to look as though no property *ever* makes a causal difference.

So, the exclusion principle dramatically overstates the potential for causal competition between properties. Not that there is nothing right about it. In *some* sense of 'separate', it stands to reason, separate properties *are* causal rivals as the principle says. Well, what if someone identifies the appropriate notion of separateness and reformulates the exclusion principle accordingly? Suppose it done. Even without hearing the details, we *know* that the corrected principle does not apply to determinates and their determinables; for we know that they are not causal rivals. Such a position is of course familiar from other contexts. Take, for instance, the claim that a space completely filled by one object can contain no other. Then are even the object's *parts* crowded out? No, in this competition parts and wholes are not on opposing teams, and any principle that puts them there needs rethinking. Likewise any credible reformulation of the exclusion principle must respect the truism that determinates and their determinables are not in causal competition.

<sup>79</sup> Depending on what exactly the exclusion principle asks in the way of causal sufficiency,  $Q'$  might be a determinate of  $Q$  only in a fairly relaxed sense. Those uncomfortable about this should remember the dialectical context: we are trying to show that the assumption needed to disempower mental properties, viz., that determinates are causally competitive with their determinables, would if true disempower virtually *all* properties. But if the assumption is true with determination strictly interpreted, then it should also be true on the looser reading; and the argument in the text now applies.



All of this goes over to particulars *mutatis mutandis*. Remember Archimedes's excited outburst on discovering the principle of displacement in his bath. That his shouting "Eureka!!" was causally sufficient (let us pretend) for the cat's startled flight cannot be thought to rule it out that his (simply) shouting was relevant as well. Equally incredible is the suggestion that, granted the causal sufficiency of Socrates's *guzzling* the poison for his death, his *drinking* it had no effect. Rather, in both these cases, as in the majority of others, the determinate's contribution *includes* the determinable's as a part. Far from being rivals, I conclude, for causal influence, determinates and determinables seem literally to share in one another's success.<sup>80</sup>

With the exclusion principle neutralised, the application to epiphenomenalism is anticlimactic. As a rule, determinates are tolerant, indeed supportive, of their determinables' causal aspirations. Why should it be different, if the determinate is physical and its determinable mental? Suppose that *P*, the physiological basis of my high spirits, was causally sufficient for my grinning. To conclude that its determinable *Q*, the property of feeling happy, was causally otiose, is no better than rejecting the redness as irrelevant on the ground that all the causal work was accomplished already by its determinate scarlet. And how could it make my pain *s* irrelevant to my wincing, that the latter was guaranteed by *s*'s occurring in some specific physical way?<sup>81</sup>

## 15. MENTAL CAUSATION

So far our position is wholly negative: for all that the exclusion argument shows, mental phenomena *can* be causally relevant compatibly with the causal sufficiency of their physical bases. It is a further question whether they *will* be in any particular case. And even if *m* is causally relevant to an effect *y*, it is a further question yet whether it actually *causes* *y*.

<sup>80</sup> I do not say that the determinable *must* be relevant if the determinate is; Yablo (1992) gives examples to the contrary.

<sup>81</sup> Suppose that causal sufficiency is read in some fairly demanding way, e.g., as requiring the nomological impossibility of *x*'s occurring without *y*'s doing so. Then no physical event *p* with hopes of determining a mental event *m* is likely to be itself causally sufficient for *m*'s presumed effect *e*. To causally *guarantee* *e*'s occurrence, *p* would need to be enormously larger than *m* in spatial terms (assuming, anyway, that *p*'s essence is not unconscionably extrinsic). But that is ruled out by *p*'s determining *m*, and their resulting coincidence. Let it be granted, then, that *p* is not causally sufficient for *e*; that honour falls instead to a spatially far more extensive physical event *p*<sup>\*</sup>, whose occurrence essentially requires, in addition to *p*, that the surrounding physical conditions should be approximately as they are in fact. This affects the question of *m*'s causal potency *only* if there is more causal rivalry between *m* and *p*<sup>\*</sup> than we found between *m* and *p* (namely, none). But, how could there be? What dispelled the illusion of rivalry between *m* and *p* was that *p*'s occurrence consisted, in part, in *m*'s occurrence, and that is as true of *m* and *p*<sup>\*</sup> as it was of *m* and *p*: for *p*<sup>\*</sup> to occur is for *m* to occur in a certain physical way, and in a certain physical environment. So *p*<sup>\*</sup> poses no greater threat than *p* to *m*'s causal ambitions.

Notice some important differences between causal relevance and sufficiency on the one hand, and causation on the other:  $x$  can be causally sufficient for  $y$  although it incorporates indefinite amounts of causally extraneous detail, and causally relevant to  $y$  even though it omits factors critical to  $y$ 's occurrence. What distinguishes causation from these other relations is that causes are expected to be *commensurate* with their effects. This makes causation special in another way also: determinables and determinates may not compete for causal *influence*, broadly conceived as including everything from relevance to sufficiency; but they *do* compete for the role of *cause*, with the more commensurate candidate prevailing. Now I argue that the effect's mental antecedents often fare *better* in this competition than their physical counterparts.

To be commensurate is, nearly enough, to be proportional. Thus faced with a choice between candidate causes, one a determinate of the other, the more proportional of the two is, other things equal, to be preferred. Which of the contenders proportionality favours depends, of course, on the effect in view. Socrates's drinking the hemlock is better positioned than his guzzling it to cause his death, but relative to other effects proportionality may back the guzzling over the drinking.

Here is an example more to the present point. In a fit of pique I decide to topple the milk pitcher. Epiphenomenalist neuroscientists are monitoring my brain activity from a remote location, and an event  $e$  in their neurometer indicates my neural condition to be thus and such. Like any mental event, my decision  $m$  has a physical determination  $p$ , and the question arises to which of these the neurometer reading  $e$  is due. The scientists reason as follows: because the neurometer is keyed to the precise condition of his brain,  $e$  would not have eventuated if the decision had been taken in a different neural way, in particular if it had occurred in  $p$ 's absence. Therefore  $m$  was not enough for  $e$ ;<sup>82</sup> and, if it was not enough, it was not  $e$ 's cause.

Before announcing this as a victory for epiphenomenalism, we should consider how things look from the interactionist's perspective. Belief in the possibility of mental causation does not entail the commitment to find it *everywhere*; and, in *this* case, no one would (should) think that the mental event was the cause. Recognising that an effect depends not just on an event  $x$ 's occurring, but on its occurring in some quite specific manner, we rightly hesitate to assign  $x$  causal credit. To treat the meter reading as resulting from my decision per se would be like attributing Zsa Zsa's citation to her driving through the police radar, or the officer's abrasions to her touching his face.

Then when do we attribute effects to mental causes? Only when we believe, I can only suppose rightly, that the effect is relatively insensitive to the finer

<sup>82</sup> Strictly speaking this assumes that *each* of  $p$ 's determinables, not just  $m$ , is such that if it had occurred in  $p$ 's absence,  $e$  would not have ensued ( $p$  can counterexample  $m$ 's claim to be enough for  $e$  only if  $e$  requires it).

details of  $m$ 's physical implementation.<sup>83</sup> Deciding to topple the pitcher, that is what I do, and the milk spills across the floor. Most people would say, and I agree, that my decision had the spill as one of its effects. As for the decision's physical determination  $p$ , most people would also say, and I agree again, that the decision would *still* have been succeeded by the spill if it had occurred in a *different* physical way (because I had taken aspirin, say, or run around the block).<sup>84</sup>

Someone could of course question this seemingly commonsensical assumption. But whoever accepts it must reckon with its *prima facie* consequences (where  $m$  is my decision,  $p$  is the brain event, and  $s$  is the spill):

- $m$  is a counterexample to  $s$ 's requiring  $p$  (for  $s$  would still have occurred, if  $m$  had occurred without  $p$ );
- $p$  is not proportional to  $s$  (since  $s$  does not require it);
- $p$  did not cause  $s$  (since it is not proportional to  $s$ );
- $p$  is not a counterexample to  $m$ 's enoughness for  $s$  (it could be a counterexample only if  $s$  required it);
- $p$  is not a counterexample to  $m$ 's proportionality with  $s$  (by inspection of the remaining conditions);
- $p$  poses no threat to the hypothesis that  $m$  caused  $s$ .

And here are the beginnings, at least, of a story in which a mental event emerges as *better* qualified than its physical basis for the role of cause.

## 16. CONCLUSION

Indeterminism aside, whatever happens is in strict causal consequence of its physical antecedents. But to be causally necessitated is a different thing from being caused, and the physical has no monopoly on causation. Among causation's prerequisites is that the cause should be commensurate with its effect; and part of commensuration is that nothing causes an effect which is essentially overladen with materials to which the effect is in no way beholden. This, though, is a condition of which would-be physical causes often fall afoul, thereby opening

<sup>83</sup> "But sometimes we want to know what is *distinctive* in an effect's etiology, i.e., how it comes about in this world as opposed to others. Then the underlying physical event might be exactly what we are after". True enough; see the discussion of world-driven causal judgements in Section 12.

<sup>84</sup> Remember that this makes no prediction about what would have happened, if the decision had occurred in *whatever* physical way, but speaks only of what happens in the *nearest* world in which the decision's physical implementation was not as actually—the world in which it undergoes only the minimal physical distortion required to put its actual implementation  $p$  out of existence. Maybe, of course, we were wrong to think that the spill would still have occurred in such a world; in that case, let us hurry to withdraw the claim that the decision caused it.

the market up to weaker events with essences better attuned to the effect's causal requirements. Sometimes, these events are mental; and that is how mental causation happens.

## REFERENCES

- Anscombe, E.: 1975, 'Causality and Determination', in Sosa (1975), pp. 63–81.
- Beauchamp, T. L. and A. Rosenberg: 1981, *Hume and the Problem of Causation*, Oxford University Press, Oxford.
- Bennett, J.: 1988, *Events and Their Names*, Hackett, Indianapolis.
- Block, N. and J. Fodor: 1972/1980, 'What Psychological States are Not', *Philosophical Review* 81, 159–81 (reprinted in Block and Fodor (1980), pp. 237–50).
- Block, N. and J. Fodor: 1980, *Readings in the Philosophy of Psychology (I)*, Cambridge University Press, Cambridge.
- Campbell, K.: 1970, *Body and Mind*, MacMillan, New York.
- Chomsky, N.: 1975, 'Remarks on Nominalization', in Davidson and Harman (1975), pp. 262–89.
- Davidson, D.: 1967/1980a, 'Causal Relations', *Journal of Philosophy* 64, 691–703 (reprinted in Davidson (1980), pp. 149–62).
- 1970/1980b, 'Mental Events', in L. Foster and J. W. Swanson (ed.), *Experience and Theory*, London, Duckworth, pp. 79–101 (reprinted in Davidson (1980), pp. 207–24).
- 1980, *Essays on Actions and Events*, Oxford University Press, Oxford.
- Davidson, D. and G. Harman: 1975, *The Logic of Grammar*, Dickenson, Encino.
- Dretske, F.: 1977, 'Referring to Events', *Midwest Studies in Philosophy* 2, 369–78.
- Dretske, F. and A. Snyder: 1973, 'Causality and Sufficiency', *Philosophy of Science* 40, 288–91.
- Ducasse, C. J.: 1969, *Causation and the Types of Necessity*, Dover, New York.
- Fodor, J.: 1987, *Psychosemantics*, MIT Press, Cambridge.
- 1989, 'Making Mind Matter More', *Philosophical Topics* 17, 59–79.
- 1991, 'A Modal Argument for Narrow Content', *Journal of Philosophy* 88, 5–26.
- Gibbard, A.: 1975, 'Contingent Identity', *Journal of Philosophical Logic* 4, 187–221.
- Goldman, A.: 1969, 'The Compatibility of Mechanism and Purpose', *Philosophical Review* 78, 468–82.
- Harper, W., R. Stalnaker, and G. Pearce (eds.): 1981, *Ifs*, D. Reidel, Boston.
- Heil, J. and A. Mele: 1991, 'Mental Causes', *American Philosophical Quarterly* 28, 49–59.
- Honderich, T.: 1982, 'The Argument for Anomalous Monism', *Analysis* 42, 59–64.
- Hume, D.: 1963, *An Enquiry Concerning Human Understanding*, Open Court, La Salle.
- 1968, *Treatise of Human Nature*, Clarendon Press, Oxford.
- Johnson, W. E.: 1964, *Logic (I)*, Dover, New York.
- Kim, J.: 1973, 'Causation, Nomic Subsumption, and the Concept of Event', *Journal of Philosophy* 70, 217–36.
- 1974, 'Non-causal Connections', *Noûs* 8, 41–52.

- Kim, J.: 1979, 'Causality, Identity, and Supervenience in the Mind-Body Problem', *Midwest Studies in Philosophy* 4, 31–50.
- 1983, 'Supervenience and Supervenient Causation', *Southern Journal of Philosophy*, supp. 22, 45–56.
- 1984a, 'Concepts of Supervenience', *Philosophy and Phenomenological Research* 45, 153–76.
- 1984b, 'Epiphenomenal and Supervenient Causation', *Midwest Studies in Philosophy* 9, 257–70.
- 1989, 'Mechanism, Purpose, and Explanatory Exclusion', *Philosophical Perspectives* 3, 77–108.
- Kripke, S. A.: 1980, *Naming and Necessity*, Harvard University Press, Cambridge.
- Lewis, D.: 1971, 'Counterparts of Persons and Their Bodies', *Journal of Philosophy* 68, 203–11.
- 1973/1986b, 'Causation', *Journal of Philosophy* 70, 556–67 (reprinted with postscript in Lewis (1986), pp. 159–72).
- 1979/1986c, 'Counterfactual Dependence and Time's Arrow', *Noûs* 13, 455–76 (reprinted in Lewis (1986), pp. 32–66).
- 1986, *Philosophical Papers (II)*, Oxford University Press, Oxford.
- 1986a, 'Events', in Lewis (1986), pp. 241–69.
- Loewer, B. and E. Lepore: 1989, 'More on Making Mind Matter More', *Philosophical Topics* 17, 175–91.
- Loewer, B. and E. Lepore: 1987, 'Mind Matters', *Journal of Philosophy* 84, 630–42.
- Lucas, J. R.: 1962, 'Causation', in R. J. Butler (ed.), *Analytical Philosophy*, Blackwell, Oxford, pp. 32–65.
- Lyon, A.: 1967, 'Causality', *British Journal for the Philosophy of Science* 18, 1–20.
- Macdonald, C. and G. Macdonald: 1986, 'Mental Causation and Explanation of Action', in L. Stevenson, L. R. Squires, and J. Haldane (eds.), *Mind, Causation, and Action*, Blackwell, Oxford, pp. 35–48.
- Mackie, J. L.: 1974, *The Cement of the Universe: A Study of Causation*, Oxford University Press, Oxford.
- Malcolm, N.: 1968/1982, 'The Conceivability of Mechanism', *Philosophical Review* 77, 45–72 (reprinted with 'Postscript' in G. Watson (ed.), *Free Will*, Oxford University Press, Oxford, pp. 127–49).
- McLaughlin, B.: 1989, 'Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical', *Philosophical Perspectives* 3, 109–35.
- Mill, J. S.: 1950, *A System of Logic*, abridged in E. Nagel (ed.), *John Stuart Mill's Philosophy of Scientific Method*, MacMillan, New York.
- Neale, S.: 1990, *Descriptions*, MIT Press, Cambridge.
- Prior, A. N.: 1949, 'Determinables, Determinates, and Determinants (I, II)', *Mind* 58, 1–20, 178–94.
- Putnam, H.: 1975a, 'On Properties', in Putnam, *Mathematics, Matter, and Method*, Cambridge University Press, Cambridge, pp. 305–22.
- 1975b, 'Philosophy and our Mental Life', in Putnam, *Mind, Language, and Reality*, Cambridge University Press, Cambridge, pp. 291–301.
- 1980, 'The Nature of Mental States', in Block and Fodor (1980), pp. 223–31.

- Quine, W. V. O.: 1953, 'Reference and Modality', in Quine, *From a Logical Point of View*, Harvard University Press, Cambridge, pp. 139–59.
- Rasmussen, S. A.: 1982, 'Ruben on Lewis and Causal Sufficiency', *Analysis* 42, 207–11.
- Russell, B.: 1912–13/1963, 'On the Notion of Cause', *Proceedings of the Aristotelian Society* 13, 1–26 (reprinted in Russell, *Mysticism & Logic*, Unwin, London, pp. 9–30).
- Schiffer, S.: 1989, *Remnants of Meaning*, MIT Press, Cambridge.
- Sorabji, R.: 1980, 'Do Coincidences Have Causes?', in *Necessity, Cause, and Blame*, Cornell University Press, Ithaca, pp. 3–25.
- Sosa, E.: 1975, *Causation and Conditionals*, Oxford University Press, Oxford.
- 1984, 'Mind-Body Interaction and Supervenient Causation', *Midwest Studies in Philosophy* 9, 271–81.
- Stalnaker, R.: 1981a, 'A Theory of Conditionals', in W. L. Harper et al. (1981), pp. 41–55.
- 1981b, 'A Defense of Conditional Excluded Middle', in Harper et al. (1981), pp. 87–104.
- Stoutland, F.: 1980, 'Oblique Causation and Reasons for Action', *Synthese* 43, 351–67.
- Taylor, R.: 1962–63/1975, 'Causation', *Monist* 47, 287–313 (reprinted in Sosa (1975), pp. 39–43).
- Thomason, R.: 1985, 'Some Issues Concerning the Interpretation of Derived and Gerundive Nominals', *Linguistics and Philosophy* 8, 73–78.
- Vendler, Z.: 1962, 'Effects, Results, and Consequences', in R. J. Butler (ed.), *Analytical Philosophy*, Barnes & Noble, New York, pp. 1–14.
- 1967/1975, 'Causal Relations', *Journal of Philosophy* 64, 704–13 (reprinted in Davidson and Harman (1975), pp. 255–61).
- Wilson, M. (ed.): 1969, *The Essential Descartes*, New American Library, New York.
- Yablo, S.: 1987, 'Identity, Essence, and Indiscernibility', *Journal of Philosophy* 84, 293–314 [Chapter 1 in this volume].
- 1992, 'Mental Causation', *Philosophical Review* 101, 245–80 (reprinted in Yablo (2009) pp. 222–48).

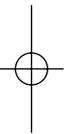


## 4

# Advertisement for a Sketch of an Outline of a Prototheory of Causation

### 1. PLATO'S DISTINCTION

A couple of thousand years before Hume made the remark that inspired the counterfactual theory of causation, Plato said something that bears on the principal problems for that theory. The idea will seem at first utterly familiar and of no possible help to anyone, so please bear with me. What Plato said, or had Socrates say, is that a distinction needs to be drawn between *the cause* and *that without which the cause would not be a cause* (*Phaedo*, 98e).



This sounds like the distinction between causes and enabling conditions: conditions that don't produce the effect themselves but create a context in which something else can do so; conditions in whose absence the something else would not have been effective. And, indeed, that is what Plato seems to have had in mind. Crito offers Socrates a chance to escape from prison. Socrates refuses and sends Crito on his way. The cause of his refusal is his judgment that one should abide by the decision of a legally constituted court. But it is facts about Socrates' body that allow the judgment to be efficacious: "if he had not had this apparatus of bones and sinews and the rest, he could not follow up on his judgment, but it remains true that it is his judgment on the question that really determines whether he will sit or run" (Taylor 1956, pp. 200–201).

Socrates' bones and sinews are factors such that if you imagine them away, the cause (Socrates' judgment) ceases to be *enough* for the effect. Are there conditions such that the cause ceases to be *required* for the effect, if you imagine them away? There seem to be. Consider an example of Hartry Field's.

BOMB: Billy puts a bomb under Suzy's chair; later, Suzy notices the bomb and flees the room; later still, Suzy has a medical checkup (it was already arranged) and receives from her doctor a glowing report.

Field intends this as a counterexample to transitivity, and so it is. The bomb is a cause of the fleeing is a cause of the glowing report; the bomb is not a





cause of the glowing report. But it is also an example of Plato's distinction. Were it not for the bomb's presence, the glowing report would not have hinged on Suzy's leaving the room. The bomb does not help Suzy's leaving to suffice for the glowing report; rather it makes Suzy's action important, required, indispensable.

Apparently there are two kinds of factors "without which the cause would not be a cause." On the one hand we have *enablers*: facts  $G$  such that  $(Oc \ \& \ \neg G) \sqsupset \neg Oe$ . On the other we have what might be called *ennoblers*: facts  $G$  such that  $(\neg Oc \ \& \ G) \sqsupset \neg Oe$ . Enablers make a dynamic contribution. They help to bring the effect about. What an ennobler contributes is just a raising of status. Suzy's removing herself from the room is elevated from something that just happens to something that *had* to happen, if Suzy was later going to be healthy.

Plato thinks that factors "without which the cause would not be a cause" are one thing, causes another. He presumably, then, would say that enablers and ennoblers are not to be regarded as causes.

About enablers, at least, it is not clear we should go along with him. If  $G$  is an enabler, then it is a fact in whose absence the effect would not have occurred. And, although there is some dispute about this, most say that that is good enough for being a cause. Enablers are full-fledged causes; it is just that they are pragmatically counterindicated in some way. Plato's distinction in its cause/enabler form can easily be rejected. Let's suppose to keep things simple that it is rejected.

Consider now ennoblers. An ennobler contributes by closing off potential routes to  $e$ , that is, all the routes not running through  $c$ . This if anything *hurts*  $e$ 's chances. So there is no question of confusing an ennobler with a cause. Plato's distinction in its (unintended) cause/ennobler form is real and important. That a potential cause is disarmed may be a factor in the manner of  $e$ 's occurrence, but it is not a factor in its occurring as such.

## 2. PREEMPTIVE CAUSES

The counterfactual theory as handed down from Hume says that "if the first did not, then the second had not been." For this to be plausible, we need to add that the first and second both occur, that the first is distinct from the second, and so on. But let us imagine that all of that is somehow taken care of, because our concerns lie elsewhere. Let the "simple" counterfactual theory be just what Hume says:

$$(CF1) \ c \text{ causes } e \text{ iff } e \text{ depends on } c, \text{ that is, } \neg O(c) \sqsupset \neg O(e)$$

The simple theory cannot be right, because it ignores the possibility of backup causes that would spring into action if the real cause failed. This is what Lewis used



to call the asymmetric overdetermination problem, and now calls preemption. An example is

DEFLECT: Hit and Miss both roll bowling balls down the lane. Hit's heavier ball deflects Miss's lighter ball en route to the pin. Hit's throw caused the pin to fall. But there is no dependence since if it had not occurred, the effect would still have happened due to a chain of events initiated by Miss's throw. (Yablo 1986, p. 143)

The problem here is obvious enough that one should probably date the counterfactual *theory*, as opposed to the immediately withdrawn counterfactual *hunch*, to the moment when it was first clearly addressed, in Lewis's 1973 paper "Causation."

Here is what Lewis says. Notice something about the chain of events initiated by Miss's throw. The Miss-chain was cut off before the Hit-chain had a chance to reach the pin. It is true that the effect does not depend on the earlier part of Hit's chain. It does, however, depend on the part occurring later, after the Miss-chain is dead and buried. And the part after the cutoff point depends in turn on Hit's throw. So the effect depends on something that depends on Hit's throw—which suggests that instead of (CF1) our analysis should be

(CF2)  $c$  causes  $e$  iff there are  $d_i$  such that  $\neg O(c) \sqsupset \neg O(d_1) \ \& \ \dots \ \& \ \neg O(d_n) \sqsupset \neg O(e)$

What (CF2) says is that causation need not be direct; it can be indirect, involving dependency chains. If the ancestral of a relation  $R$  is written  $R^*$ , it says that causation is dependence\*.

The diagnosis implicit in (CF2) is that preemption arises because we had forgotten about causal chains. I want to suggest an alternative "Platonic" diagnosis. Preemption happens because to take away a cause  $c$  is, sometimes, to take away more. It is to take away the reason it *is* a cause. It is to take away factors that, although not themselves causal, contribute to  $c$ 's causal status by putting  $e$  in need of  $c$ .  $e$  can hardly be expected to follow  $c$  out of existence, if the reasons for its depending on  $c$  disappear first.

So, look again at DEFLECT. Quoting a former self: "if in fact Miss's ball never reaches the pin, then that is an important part of the circumstances. Relative to circumstances including the fact that Miss's ball never makes it, what Hit did was necessary for the pin's toppling. If in those circumstances Hit hadn't rolled his ball down the alley, the pin would have remained standing" (Yablo 1986, p. 159). That Miss's ball never touches the pin is a fact that puts the effect in need of Hit's throw. It is a fact "in virtue of which Hit's throw is a cause." The trouble is that it is a fact put in place by the throw itself, hence one that finkishly disappears when the relation is counterfactually tested.

## 3. HOLDING FIXED

The diagnosis suggests a repair. If preemption is a matter of something finkishly giving way, the obvious thought is: Don't *let* it give way; hold the grounds of the causal connection fixed. The test of causation in these cases is not whether  $e$  fails if  $c$  does, but whether  $e$  fails if  $c$  fails *with the right things held fixed*.

By “dependence modulo  $G$ ” I will mean dependence with  $G$  held fixed. This event depends modulo  $G$  on that one iff had that one failed to occur in  $G$ -type circumstances, this one would have failed to occur as well. Letting “ $\square \rightarrow_G$ ” stand for dependence modulo  $G$ , the suggestion is that

(CF3)  $c$  causes  $e$  iff: for some appropriate  $G$ ,  $\neg O(c) \square \rightarrow_G \neg O(e)$ .

Actually, of course, this is only an analysis-schema. An analysis would require a clear, non-causation-presupposing statement of what makes for an appropriate  $G$ .

What does make for an appropriate  $G$ ? Certainly  $G$  should ennoble  $c$ . But all we have said about ennoblers is that they are conditions  $G$  such that  $e$  depends holding- $G$ -fixed on  $c$ . As you might guess, and as will be discussed below, this purely formal requirement can be met by logical trickery almost whatever  $c$  and  $e$  may be.

Thus where the standard counterfactual theory undergenerates—the events that depend on  $c$  (or depend\* on  $c$ ) are not all the events it causes—the present theory has, or is in danger of having, the opposite problem. It may well be that  $e$  depends on each of its causes modulo a suitably chosen  $G$ . But this is true also of events that do not cause  $e$ .

## 4. TRIVIALITY/POLARITY

There are actually two worries here, one building on the other. The first is a worry about trivialization; everything depends on everything modulo a silly enough  $G$ . This is illustrated by

JUMP: Suppose that  $e$  is Bob Beamon's jumping 29'2 1/2" at the 1968 Olympics in Mexico City. And let  $c$  be the burning out of a meteor many light years away. It is not hard to find facts modulo which the jump depends on the burnout. First choose an event on which the jump depends pure and simple—say, Beamon's tying his shoes. Holding it fixed that *the tying occurs only if the burnout does*, without the burnout Beamon does not make the jump.

Unless some sort of restriction is put on admissible  $G$ s, the requirement of dependence modulo  $G$  is a trivial one satisfied by any pair of events you like.

One could stipulate that  $G$  should not be too cockamamie or too ad hoc or too cooked up for the occasion. But that is hopeless. It is not just that “too ad hoc” is so vague. Suppose that a standard of naturalness is somehow agreed on. No matter how high the standard is set, there will be  $G$ -dependence without causation.

Consider again BOMB. Certainly the doctor’s glowing report does not depend simpliciter on Billy’s planting the bomb. But it does depend on it modulo the fact that *Suzy’s chair explodes*. Holding the explosion fixed, Suzy would not have been healthy unless she had moved away, which she would not have had she not noticed the bomb, which would not have been there to notice had it not been put there by Billy. It seems on the face of it insane to credit Suzy’s good health to the bomb; she is healthy despite the bomb, not because of it. And yet her health depends on the bomb modulo a natural fact. Call that the *polarity* problem.

## 5. STOCKHOLM SYNDROME

Preemptive causes make themselves indispensable. They create the conditions given which the effect would not have occurred without them. But there is more than one way of doing that. The *normal* way is to produce the effect yourself, thereby preventing other would-be causes from doing the job instead. The effect needs  $c$  modulo the fact  $G$  that other avenues to the effect are closed off.

But if you look at our basic condition—the condition of dependence modulo  $G$ —you can see that it supports an almost opposite scenario.  $E$  was going to happen anyway, when  $c$  comes along to threaten it: to put its existence in jeopardy. Of course, putting the effect in jeopardy is not all  $c$  does, or it would not even resemble a cause. It also rescues  $e$  from the jeopardy.  $C$  threatens  $e$  with one hand, and saves it with the other. The effect needs  $c$  to counter the threat  $G$  that  $c$  itself has launched.

Should  $e$  be grateful to  $c$  for blocking with one hand a threat it launches with the other? Of course not. There is a word for that kind of inappropriate gratitude. You might remember it if I quote from a Web site on the topic: “In the summer of 1973, four hostages were taken in a botched bank robbery at Kreditbanken in Stockholm, Sweden. At the end of their captivity, six days later, they actively resisted rescue. They refused to testify against their captors, raised money for their legal defense, and according to some reports one of the hostages eventually became engaged to one of her jailed captors.” Stockholm Syndrome is the gratitude hostages feel toward captors who help them with problems brought on by the captivity. To give the present sort of  $c$  causal credit would be the metaphysical equivalent of Stockholm Syndrome.

It is important to be clear about what is being rejected here. There is nothing wrong with gratitude for actions taken against a threat that has already been

launched: not even if the action is taken by the one who launched it. If your kidnapper takes pity on you and gets you a Mars bar, there is no requirement of flinging it back in his face. But suppose your kidnapper says, “I appreciate that you are grateful to me for various particular acts of mercy. But still I am hurt. Where is the thanks I get for the action that occasioned these mercies, that is, the kidnapping?” That remark you *should* fling back in his face. I draw the following moral:

Dependence modulo  $G$  does not make for causation if (i)  $G$  is a threat to  $e$  that, although (ii) countered by  $c$ , was also (iii) launched by  $c$ .

With this in mind, let’s go back to the BOMB example. There is nothing wrong with thanking the bomb for tipping you off to its presence, given that it has already been planted. But what we are talking about here is gratitude toward the planting itself. BOMB, then, is an example of Stockholm Syndrome. What may be less obvious is that JUMP is an example as well. That there will be no shoe-tying unless the meteor burns out makes  $e$  vulnerable from a new and unexpected direction; events that would cancel the burnout are now being put in a position to cancel the jump too. (That’s (i).) There would be no such fact as *shoe-tying only if burnout*, had the meteor not in fact burned out. (That’s (iii).) It is the burnout, finally, that stops this fact from carrying out its threat against the effect. (That’s (ii).)

## 6. ARTIFICIAL NEEDS

If  $e$  depends on  $c$  holding  $G$  fixed, let us say that  $G$  puts  $e$  in need of  $c$ . Why is this not enough for causation? The answer is that some needs are trumped up or artificial. This shows up in the fact that among  $e$ ’s other needs are some that would, but for  $c$ , have been *all* its needs. Or, to look at it from the point of view of the fallback scenario—the closest scenario where  $c$  does not occur— $c$  is able to meet a need only by making the effect needier than it had to be, indeed, needier than it *would* have been had  $c$  failed to occur.

Say that  $e$  is Beamon’s big jump, and  $c$  is the burning out of that meteor. What are the effect’s needs in the fallback scenario? What would the jump have depended on had the meteor not burned out? It would have depended on Beamon’s tying his shoes; on various earlier jumps whereby he won a place in the finals; on Mexico’s bid for the 1968 Olympics; and so on. *Bringing in the burnout does not diminish these needs one iota*. Everything that had to happen before, still has to happen with the meteor burning out. This is why the burnout’s role is artificial. It is strictly additional to events that meet the effect’s needs all by themselves in its absence.

Now let’s try to make this a teeny bit precise. History let’s suppose has a branching time structure. There is the trajectory actually taken through logical



space, and the various branchings-off corresponding to other ways things could have developed. One branch in particular corresponds to the way things would have developed if  $c$  had not occurred. By the *fallback scenario* let's mean what happens after the branching-off point on that alternative branch. By the *actual scenario* let's mean what actually *does* happen after the branching-off point. The effect's fallback needs are the events it depends on in the fallback scenario, that is, the events it would have depended on, had  $c$  not occurred.<sup>1</sup> Recalling our subscripting conventions from above, this can be written

$$\text{FAN} = \{ x \mid \neg O_x \sqsupset_F \neg O_e \},$$

where " $F$ " is short for " $\neg O_c$ ."<sup>2</sup> The effect's actual needs (for a given choice of  $G$ ) are the events it depends on modulo  $G$  in the actual scenario. This can be written

$$\text{GAN} = \{ x \mid \neg O_x \sqsupset_G \neg O_e \}.$$
<sup>3</sup>

The need for  $c$  is artificial iff GAN covers FAN with  $c$  to spare; or, taking the perspective of the fallback scenario, FAN is identical to a subset of GAN not including  $c$ . (I will write this  $\text{FAN} = \text{GAN}^-$ .) The point either way is that  $c$  speaks to a need that is piled arbitrarily on top of what would, in  $c$ 's absence, have been all the needs.

## 7. COUNTERPARTS

Imagine that Beamon as a child was initially attracted to chess rather than long jump. He was waiting to sign up for chess club when someone threw a rock at him. It was because of the rock incident that he wound up in track. And now here is the interesting part: the rock came from the burnt-out meteor. If not for the burnout, it would not have been *that* rock-throwing the effect depended on but a related one (in which the bully hurled a different rock). Since the rock-throwing that  $e$  would have needed is a different event from the one it does need, it would seem that GAN does not cover FAN at all, let alone with  $c$  to spare.

<sup>1</sup> If it seems odd to think of events as needs, remember that "need" can mean thing that is needed. ("The dogsled was piled high with our winter needs.") Needs in the ordinary sense do not exist in our system. Their work is done by events considered under a soon to be introduced counterpart relation, the relation of meeting-the-same-need-as.

<sup>2</sup> According to the export-import law for counterfactuals,  $A \sqsupset (B \sqsupset C)$  is equivalent to  $(A \ \& \ B) \sqsupset C$ . This implies that  $(\neg O_x \ \& \ \neg O_c) \sqsupset \neg O_e$ , the membership condition for FAN, is equivalent to  $\neg O_c \sqsupset (\neg O_x \sqsupset \neg O_e)$ , which says that  $e$  would have depended on  $x$  had  $c$  not occurred. I assume that the law is close enough to correct for our purposes, or at least that the indicated consequence is close enough to correct.

<sup>3</sup> FAN and GAN are to be understood as limited to events occurring *after* the point at which the actual world and the nearest  $c$ -less world begin to diverge.



I answer that different *event* does not have to mean different *need*. One event can meet the same need as another, as that need manifests itself in their respective scenarios.<sup>4</sup> Artificiality is still a matter of the effect's actual needs subsuming its fallback needs, so long as we understand this in the following way. Suppose that  $x$  is an event needed in the fallback scenario; one finds in GAN, not perhaps that very event, but *an* event meeting the same need (henceforth, a counterpart of  $x$ ). This complicates things a little, but not much. Where earlier we required FAN to be identical to a subset of GAN not including  $c$ , now we ask only that it coincide with such a subset, where sets coincide iff their members are counterparts.<sup>5</sup> (I will write this  $FAN \approx GAN^-$ .)

When do events speak in their respective scenarios to the same need? The idea is this. Needs that  $e$  would have had in  $c$ 's absence can be paired off with actual needs in ways that preserve salient features of the case: energy expended, distance traveled, time taken, place in the larger structure of needs. One wants to preserve as many of these features as possible, while finding matches for the largest number of needs. One asks: How much of the fallback structure is embeddable in the actual one? What is the maximal isomorphic embedding? Events speak to the same need if they are linked by this embedding.

## 8. DE FACTO DEPENDENCE

It is not enough for causation that a  $G$  can be found that puts  $e$  in need of  $c$ . Causes must meet *real* needs, and the need met by  $c$  might be trumped up or artificial. A fact  $G$  makes the need for  $c$  artificial iff it assigns other needs that would, but for  $c$ , have been all of  $e$ 's needs.

An issue I have finessed until now is how the first  $G$ —the one that puts  $e$  in need of  $c$ —lines up with the second one—the one that makes the need artificial. Suppose we say of the second  $G$  that it “enfeebles”  $c$ , as we said of the first that it ennobles it. Does it suffice for causation that an ennobler  $G$  exists that is not *itself* an enfeebler?

No, for there is almost always a  $G$  like that, namely, the material biconditional  $e$  occurs  $\leftrightarrow c$  occurs. That this ennobles  $c$  should be clear. That it does not enfeeble  $c$  can be seen as follows. (1) GAN is limited to events  $x$  on which  $c$

<sup>4</sup> Also, *same event* does not have to mean *same need*. An event that meets one need here might meet another there, or it might meet no need at all.

<sup>5</sup> I will be taking counterparthood to be symmetric and one–one. But there might be reasons for relaxing these requirements. Take first symmetry. There might be an  $x$  in FAN whose closest actual correspondent meets, not the same need as  $x$ , but a “bigger” need: one with the need met by  $x$  as a part. This closest actual correspondent ought to qualify as a counterpart of  $x$ . So, the argument goes, counterparts should be events meeting *at least* the same need, which makes counterparthood asymmetric. There is a similar worry about the one–one requirement. It might take a pair of events to meet the need  $x$  meets all by itself in the fallback scenario; or vice versa. I propose to ignore these complexities.

counterfactually depends. (If  $\neg Ox \sqsupset Oc$ ,<sup>6</sup> then  $(Oc \leftrightarrow Oe) \sqsupset (\neg Ox \sqsupset Oc)$ ; so by the export-import law,  $(\neg Ox \ \& \ (Oc \leftrightarrow Oe)) \sqsupset Oc$ ; so  $(\neg Ox \ \& \ (Oc \leftrightarrow Oe)) \sqsupset Oe$  iff  $(\neg Ox \ \& \ (Oc \leftrightarrow Oe) \ \& \ Oc) \sqsupset Oe$ ; so  $(\neg Ox \ \& \ (Oc \leftrightarrow Oe)) \sqsupset Oe$ ; so  $x$  is not in GAN.) (2) FAN is almost certainly *not* limited to events on which  $c$  counterfactually depends. That  $c$  fails to depend on  $x$  has no tendency at all to suggest that  $e$  would not have depended on  $x$  in  $c$ 's absence. (1) and (2) make it unlikely that GAN includes FAN, or hence that  $G$  enfeebles  $c$ .

Where does this leave us? It is not enough for causation that an ennobler  $G$  can be found that is not itself an enfeebler. It is, I suggest, enough that an ennobler can be found such that *no* comparably natural enfeebblers exist. And so I propose a definition

- (DD) One event *de facto depends* on another iff some  $G$  putting the first in need of the second is more natural than any  $H$  that makes the need artificial,

and I make the following claim

- (CF3)  $c$  is a *cause* of  $e$  iff  $e$  de facto depends on  $c$ .

It is understood that  $c$  and  $e$  both occur, that they are suitably distinct, and that various unnamed other conditions are met; I have in mind the same sorts of extra conditions as the counterfactual theorist uses. Sometimes (CF3) will be written (DF) to emphasize that it relies on a new type of dependence, albeit one defined in terms of counterfactual dependence.

You might have expected me to say that  $c$  is a cause iff  $e$  either depends counterfactually on  $c$  or, failing that, de facto depends on it. That formulation is fine but it is equivalent to what I did say, for de facto dependence has ordinary counterfactual dependence as a special case. If  $e$  counterfactually depends on  $c$ , then it depends on  $c$  modulo the null condition. The null condition is our ennobler, and what needs to be shown is that there are no comparably natural enfeebblers. But there cannot be enfeebblers at all, for enfeebblers presuppose fallback needs—events that  $e$  depends on in  $c$ 's absence—and  $e$  does not even occur in  $c$ 's absence.

## 9. TRIVIALITY AND POLARITY

One worry we had is that even if  $c$  and  $e$  are completely unrelated, still  $e$  is put in need of  $c$  by the fact that  $k$  occurs *only if*  $c$  occurs, where  $k$  is an event on which the effect counterfactually depends.

I say that although this is true, the victory is short-lived, because the very fact of unrelatedness means that it will be easy to find an  $H$  making the need artificial.

<sup>6</sup> I get from  $\neg(\neg Ox \sqsupset \neg Oc)$  to  $\neg Ox \sqsupset Oc$  by conditional excluded middle (CEM). CEM is generally controversial, but it seems in the present context harmless; we are not trying to show that  $Oc \leftrightarrow Oe$  is *bound* to enoble  $c$  without enfeebling it, but just that this is the likely outcome.

Usually we can let  $H$  be the null condition. That is,  $e$  counterfactually depends outright (holding nothing fixed) on events that would have been enough in  $c$ 's absence. This is just what we would expect if  $c$  has, causally speaking, nothing to do with  $e$ . Beamon's jump depends on all the same things if the burnout occurs as it would have depended on absent the burnout.

The need for  $c$  is artificial iff it is over and above what would, but for  $c$ , have been all the needs. An equivalent and perhaps clearer way of putting it is that  $c$  must either *meet* a fallback need—which it does if for some  $f$  in FAN,  $c$  meets the same need as  $f$ —or *cancel* one—which it does if for some  $f$  in FAN, no actual event meets the same need as  $f$ . The need for  $c$  is artificial iff  $c$  fails to address any fallback needs, meaning that it neither meets any fallback needs nor cancels any.

I take it as given that Billy's planting of the bomb does not meet any fallback needs. The question is whether it cancels any. Suppose that Suzy needs to stay hydrated, or she becomes very sick. She has set her Palm Pilot to remind her at noon to act on this need. The fallback scenario has her sitting quietly in her chair at noon. She has a drink of water, water being the one hydrous stuff available in the room. The actual scenario has Suzy catching her breath on the sidewalk when her Palm Pilot beeps. She eats some Italian ice, that being the one hydrous stuff available on the sidewalk. Any isomorphism worth its salt is going to associate these two events. The drinking and eating are counterparts; they speak to the same need. One imagines that the same can be done for all of the effect's fallback needs. Anything the glowing report needed absent the bomb, it still needs. The reason Billy's action is not a cause is that it fails to address any fallback needs.

Suppose that I am wrong about that. Suppose the effect's fallback needs are *not* all preserved into the actual situation; or suppose they are all preserved but one maps to the planting of the bomb. Then, I claim, the planting starts to look like a cause.

*Case 1:* There is an  $f$  in FAN such that Billy's action meets the same need as  $f$ .

Suzy needs exercise or she becomes very sick. She has set her Palm Pilot to remind her to exercise at 11:45. As things turn out, she doesn't hear the beeping because she has just spotted a bomb under her chair. Running from the bomb gives her the needed exercise and so saves her health. If that is how it goes, then Billy's planting the bomb meets the same need as would have been met by Suzy's setting her Palm Pilot. And now we are inclined to reason as follows. Billy's planting the bomb meets the need for an exercise-reminder; the need was not artificial because it would have been there bomb or not; so there is no objection to treating what Billy did as a cause.

*Case 2:* There is an  $f$  in FAN such that no actual event meets the same need as  $f$ .

Billy's planting the bomb does not in fact meet the same need as Suzy's setting her Palm Pilot. The Palm Pilot, if she had heard it, would have led Suzy to

do push-ups, thus exercising her muscles. The bomb leads her instead to run, thus exercising her heart and lungs. These are entirely different forms of exercise. Either one of them would have stopped Suzy from getting sick, but the similarity ends there. Now we are inclined to reason as follows. The effect originally had need of *muscle* exercise, that being the only kind of exercise possible in the room. It is *relieved* of that need by Billy's planting of the bomb; for Suzy now runs, thus exercising her heart and lungs. So there is no objection to treating what Billy did as a cause. (Analogy: You have a flat tire and need a jack to get back on the road. I can help you either by meeting that need, or by relieving you of it. I do the first if I provide you a jack. I do the second if I bend over and lift the car myself.)

## 10. PREEMPTION

I say that effects really do depend on their preemptive causes. There is no counterfactual dependence, because the causality rests on a fact  $G$ ; and had  $c$  not occurred, that fact would not have obtained. But we can restore the dependence by holding  $G$  fixed. I don't know how to argue for this except by going through a bunch of examples.

Recall DEFLECT. Certainly the effect is put in need of Hit's throw by the fact  $G$  that Miss's ball never gets close to the pin. It might be thought, though, that the need was artificial.

The effect's fallback needs are (let's say) for Miss's throw, her ball's rolling down the aisle, and her ball's hitting the pin. These needs would seem to recur in the actual situation as needs for Hit's throw, his ball's rolling down the aisle, his ball's hitting the pin. If that is how things line up, then Hit's throw meets the same need as was met in the fallback scenario by Miss's throw; and so the need it meets is not artificial.

Suppose on the other hand that the fallback needs are held *not* to recur in the actual situation. Then artificiality is averted through the canceling of needs rather than the meeting of them. These are intuitive considerations but they suggest that a fact that makes the need for Hit's throw artificial will not be easy to find. I do not doubt that you could construct one by brute force, but a brute force  $H$  will not be as natural as our existing  $G$ , the fact that Miss's ball never gets close.

A tradition has arisen of treating early and late preemption as very different affairs. But this is for theoretical reasons to do with Lewis's ancestral maneuver, which works for early preemption but not late; intuitively the two sorts of preemption seem much on a par. The de facto theory agrees with intuition here. Consider

DIRECT: Hit and Miss both roll balls down the lane. The balls do not come into contact. Hit's ball knocks the pin into the gutter. A moment later, Miss's ball reaches the spot where the pin formerly stood.

Once again, it is part of the circumstances that Miss's ball never gets close to the pin. That no other ball gets close puts the effect in need of Hit's throw. It is true that some  $H$  might expose the need as artificial. But such an  $H$  would have to be constructed by brute force. There is no more reason to expect a natural enfeebler in this case than in the previous one.<sup>7</sup>

## 11. OVERDETERMINATION

Overdetermination occurs when an effect  $e$  depends on two events taken together without depending on either taken alone; and (what distinguishes it from preemption) neither can lay claim to being more of a cause than the other. Consider

TOGETHER: Knock and Smack roll their balls at the same time; the balls hit the pin together and it falls over; either ball alone would have been enough.

It is not hard to find suitable  $G$ s. The effect depends on Knock's throw, holding fixed the fact  $G_k$  that Smack's ball does not hit the pin unaccompanied, that is, unless another ball also hits. And it depends on Smack's throw, holding fixed the fact  $G_s$  that Knock's ball does not hit the pin unaccompanied.

It is not hard to find suitable  $H$ s either; indeed we have already found them.  $G_k$  makes the need for Smack's throw artificial, and  $G_s$  does the same for Knock's. To see why, suppose that Knock had not thrown. The effect would have depended on Smack's throw, the forward motion of his ball, and the like. These events are still needed in the actual situation, if we hold fixed the fact  $G_s$  that Knock's ball does not hit the pin alone.

Assuming that these are the most natural cause-makers and -breakers to be had, does the de facto theory call Knock's throw (e.g.) a cause? Is the effect put in need of it by a fact more natural than any fact making the need artificial?

That depends. One reading of "more natural" is *strictly* more natural. If that is what is meant, then neither throw is a cause; each prima facie connection is broken by a fact exactly as natural as the one that established it. But the phrase could also be taken weakly, to mean "at least as natural as." If, as claimed, the makers and breakers are the same, then the weak reading makes both throws out to be causes. True, each occurs under conditions given which the effect takes

<sup>7</sup> What if we change the example so that Miss's ball *does* hit the pin, after it has been knocked down? Then  $G$  should be this: Miss's ball never gets close to the pin when it is in an upright position, i.e., when it is in a condition to be toppled. Holding fixed that Miss's ball never approaches the pin at any *relevant* time, it remains the case that without Hit's throw, the pin would not have been knocked over.

no notice of it; but then each also occurs under conditions no less natural given which the effect needs it. Ties go to the runner on the weak reading, so we have two bona fide causes. Our uncertainty about overdeterminers reflects indecision about what to mean by “more.” (This is intended less as an explanation of the uncertainty than a rational reconstruction of it.)

## 12. ASYMMETRY

Suppose that  $c$  affects not whether  $e$  occurs but only when it occurs. Could that be enough to make  $c$  a cause? An example is given by Jonathan Bennett.

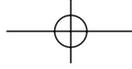
RAINDELAY: “There was heavy rain in April and electrical storms in the following two months; and in June the lightning took hold and started a forest fire. If it hadn’t been for the heavy rain in April, the forest would have caught fire in May” (Bennett 1987, p. 373).

Bennett says that “no theory should persuade us that delaying a forest’s burning for a month (or indeed a minute) is causing a forest fire. . . .” And then he points out something interesting. “Although you cannot cause a fire by delaying something’s burning, you can cause a fire by hastening something’s burning” (ibid.). So, consider

LIGHTNING: There are no rains in April. The fire happens in May owing to May lightning, rather than in June owing to the lightning that strikes then. The lightning is a cause of the fire even though the fire would still have occurred without it. That the time of occurrence would have been later rather than earlier seems to make all the difference.

Bennett’s examples raise two problems for standard counterfactual accounts. One is that they cannot explain the *asymmetry*, that is, why hasteners seem more like causes than delayers. Also, though, they have trouble explaining *why there should be causation here at all*. I assume with Bennett that hasteners bring it about that the very same event occurs earlier than it would have. If in fact the fire would still have occurred without the lightning, how can the lightning be regarded as a cause?

The form of that question ought to seem pretty familiar. It is the standard preemption question: How can  $c$  be a cause, when the effect would have occurred without it thanks to  $c'$  waiting in the wings? The answer is the same as always: It is a cause because the effect depends on it modulo a certain fairly natural fact, and nothing that natural exposes the dependence as fraudulent. It is a part of the circumstances that *the woods do not catch fire in June (or later)*. Holding that fixed, without the May lightning there would not have been a fire. The May lightning causes the fire because the fire depends on it, holding fixed that May is its last opportunity.



But there is an obvious objection. The effect also fails to occur *before* a certain time, and this would seem to obliterate the intended asymmetry. Holding fixed the lack of a fire before June, if not for April's rain there would not have been a fire at all. June was the window of possibility, and it was the rain that kept the forest going until that window opened.

The difference between rain and lightning is not that the first meets no need; rather, it has to do with the kind of need involved. Suppose the rain had not fallen, so that the forest burned in May. Then the things that were done to preserve it from May until June would not have been required. (The loggers wouldn't have had to go on strike, the rangers wouldn't have had to apply the flame retardant, and so on.) That the rain introduces new needs would not be a problem if it addressed some old ones. But it doesn't. The things that would have been needed for the May fire, had the rain not fallen, continue to be needed as conditions of the June fire. (A landslide late in April threatens to bury the forest under rubble; the June fire needs it to change course just as much as the May fire would have.)

Now we see why the rain makes a bad cause. It piles on new needs without canceling any old ones. The lightning, by contrast, cancels a whole month of old needs. The pattern here is typical of the genre. Just by their definition, hasteners are liable to speak to fallback needs; they reduce the time period over which the effect is in jeopardy and so cancel any needs pertaining to the period that is chopped off. Just by their definition, delayers often bring about a situation in which the effect needs more than it would have had the delaying event not occurred. The effect is in jeopardy for longer and has needs pertaining to the extra time. This is why hasteners tend to be causes and delayers tend not to be.

### 13. THE HASTENER THEORY

I have treated hastening as a special case of preemption. One might try the reverse, assimilating preempters to hasteners (Paul 1998b). A cause is an event in whose absence  $e$  would not have occurred, or at any rate would not have occurred as early as it did. If we count never occurring as the limiting case of delay, then the claim is that causes are hasteners, that is, events in whose absence the effect would have been delayed. One problem for this view is that hasteners are not always causes. Here is an example due to Hugh Rice (1999, p. 160):

REFLEXES: Slow Joe and Quick-Draw McGraw are shooting at Billy the Kid. Joe fires first, but since his gun fires slower-moving bullets, it is not too late for McGraw (if he fires) to cause the death. And so it happens. "McGraw (blest . . . with super fast reflexes) was aware of Joe's firing and as a result (wishing to have the glory of killing Billy for himself) fired a little



earlier than he would otherwise have done. . . . It seems that McGraw's firing was a cause of  $e$ , but that Joe's firing was not" (ibid.). Both shots hasten the death. So both count on the hastener theory as causes. Intuitively, however, it is McGraw's shot that kills Billy.

What does the de facto theory say about this? It is not hard to find a  $G$  modulo which the death depends on McGraw's shot. As the situation in fact develops, Joe's bullet never comes into contact with Billy (it passes untouched through the hole left by McGraw's bullet). Holding that fixed, Billy's death would not have occurred were it not for McGraw. This same  $G$  also enfeebles Joe's shot. Had Joe not fired, Billy's death would have depended on McGraw's shot, the motion of his bullet, and so on. Those are its fallback needs. The death's actual needs are the events on which it depends holding fixed that Joe's bullet never made contact. Prima facie it would seem that the death's fallback needs are all preserved into the actual scenario: Anything the effect depended on absent Joe's shot, it continues to depend on given that Joe's bullet doesn't hit anything.

I said that hasteners tend to reduce needs pertaining to the time period over which the effect is no longer in jeopardy. That assumes, however, that the counterpart relation puts a lot of emphasis on temporal as opposed to other factors. Oftentimes other factors will seem just as important, or more important. Suppose that by kicking a bowling ball already en route to the pin, I get it to arrive more quickly. Ordinarily my kick would count as a cause. This time, though, the main threat to the ball's forward motion is from equally spaced gates that open and shut according to a complicated pattern. The effect occurs only if the ball makes it through each of the gates. Then we might feel that the effect's needs are better conceptualized in terms of number of gates than number of seconds. To the extent that kicking the ball leaves its chances with the gates unchanged, the "need" it meets comes to seem artificial. Certainly the kick seems like less of a cause when it is stipulated that the obstacles are distributed spatially rather than temporally.

I said that delayers often bring about a situation in which the effect needs strictly more than it would have, had the delaying event not occurred. The effect is in jeopardy for longer and has needs pertaining to the extra time. But again, this is only a trend, not a strict rule. Sometimes by putting an effect off for a bit we can cut down on other, more important needs. Consider a variant of REFLEXES: McGraw is standing further from Billy than Slow Joe. When Joe sees that McGraw has fired, he fires his slower bullet on a trajectory that has it deflecting McGraw's bullet off to the side before reaching Billy. Joe's firing makes the effect happen later than it would have, but it is still a cause. Counterparthood is judged in respect not of time but of dependency relations; Joe's firing meets the need that McGraw's would have met, or, on an alternative accounting, it cancels it. It is Joe's shot that kills Billy, despite the fact that Billy lives a little longer because of it.

## 14. TRUMPING PREEMPTION

A second recent response to the preemption problem focuses on events causally intermediate between  $c$  and  $e$ . It exploits the fact that, in all the usual cases,  $e$  would have depended on events other than those actual intermediaries had  $c$  failed to occur (Ganeri, Noordhof, and Ramachandran 1998). A third focuses on the manner in which the effect occurs, if caused by something other than  $c$ . There is nothing in the nature of preemption, though, that requires intermediate events, or that the effect's characteristics should vary according to its cause.

SPELL: Imagine that it is a law of magic that the first spell cast on a given day [matches] the enchantment that midnight. Suppose that at noon Merlin casts a spell (the first that day) to turn the prince into a frog, that at 6:00 P.M. Morgana casts a spell (the only other that day) to turn the prince into a frog, and that at midnight the prince becomes a frog. Clearly, Merlin's spell . . . is a cause of the prince's becoming a frog and Morgana's is not, because the laws say that the first spells are the consequential ones. Nevertheless, there is no counterfactual dependence of the prince's becoming a frog on Merlin's spell, because Morgana's spell is a dependency-breaking backup. Further, there is neither a failure of intermediary events along the Morgana process (we may dramatize this by stipulating that spells work directly, without any intermediaries), nor any would-be difference in time or manner of the effect absent Merlin's spell. . . . Thus nothing remains by which extant [counterfactual accounts of causation] might distinguish Merlin's spell from Morgana's in causal status. (Schaffer, 2000, p. 165)

What does our sketch of a prototheory say about this case? First, we should look for a  $G$  such that the effect depends modulo  $G$  on Merlin's spell. How about the fact that no one casts a spell before Merlin does? Holding that fixed, there would have been no transformation had Merlin not cast his spell. Can a no less natural  $H$  be found that enfeebles Merlin's spell? I have not been able to think of one. It is perhaps enough to show that, unlike the other approaches mentioned, the de facto dependence account is not at an absolute loss here.

## 15. SWITCHING

A switch is an event that changes the route taken to the effect. It may not be obvious how switching so described goes beyond standard preemption, but consider an example.

YANK: A trolley is bearing down on a stalled automobile. The car lies 110 yards ahead on the track—or rather tracks, for just ahead the track splits

into two 100-yard subtracks that reconverge ten yards short of the car. Which subtrack the trolley takes is controlled by the position of a switch. With the switch in its present position, the trolley will reach the car via subtrack *U* (for unoccupied). But Suzy gives the switch a yank so that the trolley is diverted to subtrack *O* (for occupied). It takes subtrack *O* to the reconvergence point and then crashes into the car.

Certainly the crash does not counterfactually depend on the yank; had Suzy left the switch alone, the trolley would have taken subtrack *U* to the car, and the crash would have occurred as ever. Thus the simple counterfactual theory (CF1) does not classify the yank as a cause. The ancestralized theory (CF2) sees things differently; the effect depends on events that depend on the yank—the trolley’s regaining the main line from track *O*, for instance—so what Suzy did was a cause. (The verdict does not change if track *O* was mined; Suzy was hoping to get the trolley blown up, and would have succeeded had the bomb squad not arrived.)

What does the de facto theory say? There is no trouble finding a *G* such that the crash depends modulo *G* on the yank. Holding fixed that subtrack *U* is untraveled, had the switch not been pulled there would have been no way forward; the trolley would, let’s assume, have derailed. The worry is that some comparably natural *H* makes the need artificial. And, indeed, the null fact makes it artificial. Here in the actual scenario, the effect has need of 100 one-yard motions down track *O*. Had the yank not occurred, its needs would have been for 100 one-yard motions down track *U*. Because the yank lies apart from what might as well have been all the effect’s needs, the de facto theory does not call it a cause.

The de facto theory lets the yank be a cause iff it either *meets* a fallback need or *cancel*s one. As the case was first stated, it does neither, but suppose we tweak it a little. Suppose that *O* is shorter, or that *U* was not connected until after Suzy pulled the switch. Then there are needs the effect would have had that the yank does away with, and so the role it plays is not entirely artificial. Alternatively, suppose the switch operates not by rearranging the tracks, but by physically grabbing hold of the train and forcing it away from *U* and down *O*. Then the yank does meet a fallback need, the one that would in its absence have been met by the train’s continued momentum. The door is thus open to the yank’s being classified as a cause.

This is a good place to acknowledge that although *technically*, everything that *e* would have depended on counts as a fallback need,<sup>8</sup> *in practice* not all such needs are taken equally seriously. Suppose that track *U* has been disconnected for years, and heroic efforts are required to fix it. That it makes those efforts unnecessary earns the yank causal credit. But what if the track is constantly reversing itself; it is part of *U*’s design to connect when it senses an approaching trolley and

<sup>8</sup> Remember that attention is limited to events occurring *after* the “branch-point”: the point at which the nearest *c*-less world begins to depart from actuality.

disconnect when the trolley is safely past. Then the need that gets canceled may be considered too slight to protect the yank from charges of artificiality. I have no criterion to offer of when a fallback need is sufficiently serious that  $c$  can escape artificiality by canceling it. But two relevant questions are these: were the effect to fail, what are the chances of its failing for lack of  $x$ ? And how counterfactually remote are the scenarios where  $x$  is the culprit? A fallback need may not count for much if it is the last thing one would think of as the reason why  $e$  would fail.

Some have said that an event that makes “minor” changes in the process leading to  $e$  is not its cause, whereas an event that makes “major” changes is one. Our theory agrees, if “minor” changes are changes whereby all the same needs have to be met. Consider in this connection an example of Ned Hall’s, in “Causation and the Price of Transitivity” (2000, p. 209):

THE KISS: One day, [Billy and Suzy] meet for coffee. Instead of greeting Billy with her usual formal handshake, however, Suzy embraces him and kisses him passionately, confessing that she is in love with him. Billy is thrilled—for he has long been secretly in love with Suzy, as well. Much later, as he is giddily walking home, he whistles a certain tune. What would have happened had she not kissed him? Well, they would have had their usual pleasant coffee together, and afterward Billy would have taken care of various errands, and it just so happens that in one of the stores he would have visited, he would have heard that very tune, and it would have stuck in his head, and consequently he would have whistled it on his way home. . . . But even though there is the failure of counterfactual dependence typical of switching cases (if Suzy hadn’t kissed Billy, he still would have whistled), there is of course no question whatsoever that as things stand, the kiss is among the causes of the whistling.

That seems right: The kiss is among the causes of the whistling. But the example is not really typical of switching cases, or at least, it is missing features present in “pure” cases like YANK. The effect’s fallback needs (its needs absent the kiss) are heavily weighted toward the period after Billy leaves the coffee shop. They include, for instance, Billy’s deciding to drop into that particular store, the store’s staying open until he arrives, the playing of that particular tune, and so on. It is because Suzy’s kiss relieves the effect of this heavy burden of late-afternoon needs that we are ready to accept it as a cause.

## REFERENCES

- Bennett, Jonathan (1987). “Event Causation: The Counterfactual Analysis”. *Philosophical Perspectives*, vol. 1 (Ridgeview, Atascadero), pp. 367–86.
- Ganeri, J., Noordhof, P., and Ramachandran, M. (1998), “For a (Revised) PCA-Analysis”. *Analysis*, 58(1): 45–7.

- Hall, Ned (2000). "Causation and the Price of Transitivity". *Journal of Philosophy* 97(4): 198–222.
- Lewis, David (1973). "Causation". *Journal of Philosophy*, 70(4): 556–7.
- Paul, Laurie A. (1998). "Keeping Track of the Time: Emending the Counterfactual Analysis of Causation". *Analysis*, 58(3): 191–8.
- Rice, Hugh (1999). "David Lewis's Awkward Cases of Redundant Causation". *Analysis*, 59(3): 157–64.
- Schaffer, Jonathan (2000). "Trumping Preemption". *Journal of Philosophy* 99(3): 130–148.
- Taylor, A. E. (1956), *Plato: The Man and His Work* (New York: Meridian).
- Yablo, Stephen (1986), *Things*. Dissertation. University of California, Berkeley.
- (2002). "Defacto dependence". *Journal of Philosophy* 99(3): 130–148. ≡



## 5

### Does Ontology Rest on a Mistake?

Not that I would undertake to limit my use of the words ‘attribute’ and ‘relation’ to contexts that are excused by the possibility of such paraphrase . . . consider how I have persisted in my vernacular use of ‘meaning’, ‘idea’, and the like, long after casting doubt on their supposed objects. True, the use of a term can sometimes be reconciled with rejection of its objects; but I go on using the terms without even sketching any such reconciliation.<sup>1</sup>

Quine, *Word and Object*

#### I



*Introduction.* Ontology the progressive research program (not to be confused with ontology the swapping of hunches about what exists) is usually traced back to Quine’s 1948 paper ‘On What There Is’. According to Quine in that paper, the ontological problem can be stated in three words—‘what is there?’—and answered in one: ‘everything’. Not only that, Quine says, but ‘everyone will accept this answer as true’.

If Quine is right that the ontological problem has an agreed-on answer, then what excuse is there for a subject called ontology?

Quine’s own view on this comes in the very next sentence: ‘there remains room for disagreement over cases’. Of course, we know or can guess the kind of disagreement Quine is talking about.<sup>2</sup> Are there or are there not such entities as the number nineteen, the property of roundness, the chance that it will rain, the month of April, the city of Chicago, and the language Spanish? Do ‘they’ really exist or do we have here just grammar-induced illusions?

And yet, there is a certain cast of mind that has trouble taking questions like these seriously. Some would call it the *natural* cast of mind: it takes a good deal

I owe thanks to David Velleman, Ken Walton, Jamie Tappenden, Marc Kelly, Eunice Lee, Jacob Howard, George Wilson, Susan Wolf, Bas van Fraassen, Laura Bugge Schroeter, and especially David Hills. Some sections of this chapter are also reproduced in Chapter 6.

<sup>1</sup> Quine 1960, 210.

<sup>2</sup> Quine 1960 lists ‘disagreement on whether there are wombats, unicorns, angels, neutrinos, classes, points, miles, propositions’ (233).



of training before one can bring oneself to believe in an undiscovered fact of the matter as to the existence of nineteen, never mind Chicago and Spanish. And even after the training, one feels just a teensy bit ridiculous pondering the ontological status of these things.

Quine of course takes existence questions dead seriously.<sup>3</sup> He even outlines a program for their resolution: Look for the best overall theory—best by ordinary scientific standards or principled extensions thereof—and then consider what has to exist for the theory to be true.

Not everyone likes this program of Quine's. Such opposition as there has been, though, has centred less on its goals than on technical problems with the proposed method. Suppose a best theory were found; why shouldn't there be various ontologies all equally capable of conferring truth on it? Isn't a good theory in part an ontologically plausible one, making the approach circular?<sup>4</sup>

But again, there is a certain cast of mind that balks rather at the program's goals. A line of research aimed at determining whether Chicago, April, Spanish, etc. really exist strikes this cast of mind as naive to the point of comicality. It's as though one were to call for research into whether April is really the cruellest month, or Chicago the city with the big shoulders, or Spanish the loving tongue. (The analogy is not entirely frivolous as we will see.)

## II

*Curious/Quizzical.* Here then are two possible attitudes about philosophical existence-questions: the *curious*, the one that wants to find the answers, and the *quizzical*, the one that doubts there is anything to find and is inclined to shrug the question off.

Among analytic philosophers the dominant attitude is one of curiosity.<sup>5</sup> Not only do writers on numbers, worlds, and so on give the impression of trying to work out whether these entities are in fact there, they almost always adopt Quine's methodology as well. An example is the debate about sets. One side maintains with Putnam and Quine that the indispensability of sets in science argues for their reality; the other side holds with Field and perhaps Lewis that sets are not indispensable and (so) can safely be denied. Either way, the point is to satisfy curiosity about what there is.

<sup>3</sup> I am talking about the 'popular', pre-late-1960s, Quine: the one who wrote 'A logistical approach to the ontological problem', 'On what there is' (ignoring the ontological relativism), 'Two dogmas of empiricism', 'On Carnap's views on ontology', and *Word & Object* (ignoring the ontological relativity). Quine's later writings are not discussed here at all.

<sup>4</sup> Doubts have been expressed too about the extensionality of Quinean commitment. Particularly helpful on these topics are Chomsky & Scheffler 1958–9, Stevenson 1976, and Jackson 1980.

<sup>5</sup> It might be safer to say that curiosity is the analytic movement's 'official' attitude, the one that most published research unapologetically presupposes. (This after a period of ordinary-language-inspired quizzicality, as in Ryle 1954, 'The World of Science and the Everyday World'.)

How many philosophers lean the other way is not easy to say, because the quizzical camp has been keeping a low profile of late. I can think of two reasons for this, one principled and the other historical.

The principled reason is that no matter how oddly particular existence-claims, like ‘Chicago exists’, may fall on the ear, existence as such seems the very paradigm of an issue that has to admit of a determinate resolution. Compare in this respect questions about *whether* things are with questions about *how* they are.

How a thing is, what characteristics it has, can be moot due to features of the descriptive apparatus we bring to bear on it. If someone wants to know whether France is hexagonal, smoking is a dirty habit, or the Liar sentence is untrue, the answer is that no simple answer is possible. This causes little concern because there’s a story to be told about why not; the predicates involved have vague, shifty, impredicative, or otherwise unstraightforward conditions of application.

But what could prevent there from being a fact of the matter as to *whether* a thing is? The idea of looking for trouble in the application conditions of ‘exists’ makes no sense, because these conditions are automatically satisfied by whatever they are tested against.

Don’t get me wrong; the feeling of mootness and pointlessness that some existence-questions arouse in us is a real phenomenological datum that it would be wrong to ignore. But a feeling is, well, only a feeling. It counts for little without a *vindicating explanation* that exhibits the feeling as worthy of philosophical respect. And it is unclear how the explanation would go, or how it could possibly win out over the non-vindicating explanation that says that philosophical existence-questions are just very hard.

This connects up with the second reason why the quizzical camp has not been much heard from lately. The closest thing the quizzicals have had to a champion lately is Rudolf Carnap in ‘Empiricism, Semantics, and Ontology’. This is because Carnap *had* a vindicating explanation to offer of the pointless feeling: The reason it feels pointless to ponder whether, say, numbers exist is that ‘numbers exist’, as intended by the philosopher, has no meaning.<sup>6</sup> Determined to pronounce from a position external to the number-framework, all the philosopher achieves is to cut himself off from the rules governing the use of ‘number’, which then drains his pronouncements of all significance.

Quine’s famous reply (see below) is that the internal/external distinction is in deep cahoots with the analytic/synthetic distinction and just as misconceived. That Carnap is widely seen to have *lost* the ensuing debate is a fact from which the quizzical camp has never quite recovered. Carnap’s defeat was

<sup>6</sup> So says my Carnap, anyway; for a sense of the interpretive options see Haack, Stroud, Hookway, and Bird.

indeed a double blow. Apart from embarrassing the quizzicals' champion, it destroyed the only available model of how quizzicalism might be philosophically justified.

### III

*Preview.* I don't especially want to argue with the assessment of Carnap as loser of his debate with Quine. Internal/external<sup>7</sup> as Carnap explains it *does* depend on analytic/synthetic. But I think that it can be freed of this dependence, and that once freed it becomes something independently interesting: the distinction between statements made within make-believe games and those made without them—or, rather, a special case of it with some claim to be called the metaphorical/literal distinction.

This make-believish twist turns the tables somewhat. Not even Quine considers it ontologically committing to say in a *figurative* vein that there are *Xs*. His program for ontology thus presupposes a distinction in the same ballpark as the one he rejects in Carnap. And he needs the distinction to be tolerably clear and sharp; otherwise there will be no way of implementing the exemption from commitment that he grants to the non-literal.

Now, say what you like about analytic/synthetic, compared to the literal/metaphorical distinction it is a marvel of philosophical clarity and precision. Even those with use for the notion admit that the boundaries of the literal are about as blurry as they could be, the clear cases on either side enclosing a vast interior region of indeterminacy.

An argument can thus be made that it is Quine's side of the debate, not Carnap's, that is invested in an overblown distinction. It goes like this: To determine our commitments, we need to be able to ferret out all traces of non-literality in our assertions. If there is no feasible project of doing *that*, then there is no feasible project of Quinean ontology. There may be quicker ways of developing this objection, but the approach through 'Empiricism, Semantics, and Ontology' is rich enough in historical ironies to be worth the trip.

### IV

*Carnap's proposal.* Existence-claims are not singled out for special treatment by Carnap; he asks only that they meet a standard to which all meaningful talk is subject, an appropriate sort of discipline or rule-governedness. Run through his formal theory of language, this comes to the requirement that meaningful

<sup>7</sup> 'Internal/external' is short for 'the internal/external distinction'; likewise 'analytic/synthetic'.

discussion of *Xs*—material objects, numbers, properties, spacetime points, or whatever—has got to proceed under the auspices of a *linguistic framework*, which lays down the ‘rules for forming statements [about *Xs*] and for testing, accepting, or rejecting them’.<sup>8</sup> An ontologist who respects this requirement by querying ‘the existence of [*Xs*] *within the framework*’ is said by Carnap to be raising an *internal* existence-question.<sup>9</sup>

A good although not foolproof way to recognize internal existence-questions is that they tend to concern, not the *Xs* as a class, but the *Xs* meeting some further condition: ‘is there a piece of paper on my desk?’ rather than ‘are there material objects?’ I say ‘not foolproof’ because one *could* ask in an internal vein about the *Xs* generally; are there these entities or not? The question is an unlikely one because for any framework of interest, the answer is certain to be ‘yes’. (What use would the *X*-framework be if having adopted it, you found yourself with no *Xs* to talk about?) But both forms of internal question are possible.

The point about internal existence-questions of either sort is that they raise no difficulties of principle. It is just a matter of whether applicable rules authorize you to say that there are *Xs*, or *Xs* of some particular kind. If they do, the answer is *yes*; otherwise *no*; end of story.<sup>10</sup> This alone shows that the internal existence-question is not the one the philosopher meant to be asking: it is not the ‘question of realism’. A system of rules making ‘there are material objects’ or ‘there are numbers’ *unproblematically* assertible is a system of rules in need of external validation, or the opposite. Are the rules right to counsel acceptance of ‘there are *Xs*’? It is no good consulting the framework for the answer; we know what *it* says. No, the existence of *Xs* will have to be queried from a position outside the *X*-framework. The philosopher’s question is an *external* question.

Now, Carnap respects the ambition to cast judgment on the framework from without. He just thinks philosophers have a wrong idea of what is coherently possible here. How can an external deployment of ‘there are *Xs*’ mean anything, when by definition it floats free of the rules whence alone meaning comes?

There are of course meaningful questions in the vicinity. But these are questions that mention ‘*X*’ rather than using it: e.g., the practical question ‘should we adopt a framework requiring us to use “*X*” like so?’<sup>11</sup> If the philosopher protests that she meant to be asking a question about *Xs*, not the term ‘*X*’, Carnap has a ready reply: ‘You also thought to be asking a meaningful question, and one external to the *X*-framework. And it turns out that these conditions cannot be reconciled. The best I can do by way of indulging your desire to query the framework itself is to hear you as asking a question of advisability’.

<sup>8</sup> Carnap 1956, 208.

<sup>9</sup> Carnap 1956, 206.

<sup>10</sup> I am slurring over the possibility that the rules yield no verdict; cf. the treatment of solubility judgments in Carnap 1936/7.

<sup>11</sup> Also mentioned is the theoretical question, ‘how well would adopting this framework serve our interests as inquirers?’.

So that is what he does; the ‘external question’ becomes the practical question, and the ‘question of realism’ which the philosopher thought to be asking is renounced as impossible. There is something that the ‘question of realism’ was *supposed* to be; there is a concept of the question, if you like. But the concept has no instances.<sup>12</sup>

## V

*Internal/external and the dogma of reductionism.* Quine has a triple-barrelled response, set out in the next three sections.<sup>13</sup> The key to Carnap’s position (as he sees it) is that ‘the statements commonly thought of as ontological are proper matters of contention only in the form of linguistic proposals’.<sup>14</sup> But now, similar claims have been made about the statements commonly thought of as *analytic*; theoretical-sounding disputes about whether, say, the square root of  $-1$  is a number are best understood as practical disputes about how to use ‘number’. So, *idea*: the external existence-claims can be (re)conceived as the analytic ones. The objection thus looks to be one of guilt-by-association-with-the-first-dogma: ‘if there is no proper distinction between analytic and synthetic, then no basis at all remains for the contrast which Carnap urges between ontological statements and empirical statements of existence’.<sup>15</sup>

Trouble is, the association thus elaborated doesn’t look all that close. For one thing, existence-claims of the kind Carnap would call analytic show no particular tendency to be external. Quine appreciates this but pronounces himself unbothered: ‘there is in these terms no contrast between analytic statements of an ontological kind and other analytic statements of existence such as “There are prime numbers above a hundred”; but I don’t see why he should care about this’.<sup>16</sup> Quine’s proposal also deviates from Carnap in the opposite way; existence-claims can fail to be analytic without (on that account) failing to

<sup>12</sup> Is the concept incoherent? On my interpretation, yes. Yet as Bird remarks, Carnap says only that the question of realism has not been made out. I read the relevant passages as leaving the door open, not to the question of realism as he defines it (*his* definition can’t be satisfied), but to an alternative definition.

<sup>13</sup> Quine devotes most of his 1951b to another, seemingly much sillier, objection. See Bird for criticism.

<sup>14</sup> Quine 1951b, 71.      <sup>15</sup> Quine 1951b, 71.

<sup>16</sup> What is so hard to see? Internal/external was supposed to shed light on the felt difference between substantive, ‘real world’, existence-questions and those of the sort that only a philosopher could take seriously. ‘Are there primes over a hundred?’ as normally understood falls on one side of this line; ‘are there numbers?’ as normally understood falls on the other. Carnap should thus care very much if Quine’s version of his distinction groups these questions together. The problem is by no means an isolated one. According to Carnap in the Schilpp volume, existence-claims about abstract objects are ‘*usually* analytic and trivial’ (Schilpp 1963, 871, emphasis added).

be external. An example that Carnap himself might give is ‘there are material objects’. Quine apparently considers it a foregone conclusion that experience should take a course given which ‘there are material objects’ is assertible in the thing framework.<sup>17</sup> How could it be? It is not analytic that experience even occurs.<sup>18</sup>

All of that having been said, Carnap agrees that the distinctions are linked: ‘Quine does not acknowledge [my internal/external] distinction’ because according to him ‘there are no sharp boundary lines between logical and factual truth, questions of meaning and questions of fact, between acceptance of a language structure and the acceptance of an assertion formulated in the language’.<sup>19</sup> The parallel here between ‘logical truth’, ‘questions of meaning’, and ‘acceptance of a language structure’ suggests that analytic/synthetic may define internal/external (not directly, by providing an outright equivalent, but) *indirectly* through its role in the notion of a framework. The assertion rules that make up frameworks are not statements, and so there is no question of calling them analytically *true*. But they are the nearest thing to, namely, analytically *valid* or *correct*. The rules are what give *X*-sentences their meanings, hence they ‘cannot be wrong’ as long as those meanings hold fixed.

Pulling these threads together, internal/external presupposes analytic/synthetic by presupposing frameworkhood; for frameworks are made up inter alia of analytic assertion rules. Some might ask, ‘why should analytic rules be as objectionable as analytic truths?’ But that is essentially to ask why Quine’s second dogma—the reductionism that finds every statement to be linkable by fixed correspondence rules to a determinate range of confirming observations—should be as objectionable to him as the first. The objection is the same in both cases. Any observation can work for or against any statement in the right doctrinal/methodological context. Hence no assertion *or rule of assertion* can lay claim to being indefeasibly correct, as it would have to be were it correct as a matter of meaning. Quine may be right that the two dogmas are at bottom one; still, our finding *narrowly* drawn is one of guilt-by-association-with-the-second-dogma.

<sup>17</sup> He includes it on a list of sentences said to be ‘analytic or contradictory given the language’ (Quine 1951b, 71). Why a true-in-virtue-of-meaning sentence would be well suited for the role of a sentence that is untrue-in-virtue-of-being-cognitively-meaningless is not altogether clear.

<sup>18</sup> On the other hand: ‘Accepting a new kind of entity’ involves, for Carnap, adopting a new style of variable with corresponding general term. ‘There are material objects’ thus translates as  $(\exists m) \text{MATOBJ}(m)$ ; which, given how the variable and term are coordinated, is equivalent to  $(\exists m)m=m$ ; which, to come at last to the point, is logically valid in standard quantificational logic. On the third hand, Carnap *objected* to this feature of standard quantificational logic: ‘If logic is to be independent of empirical knowledge, then it must assume nothing concerning the *existence of objects*’ (Carnap 1937, 140). In his ‘physical language’, he notes, ‘whether anything at all exists—that is to say, whether there is . . . a non-trivially occupied position—can only be expressed by means of a synthetic sentence’ (ibid., 141).

<sup>19</sup> Carnap 1956, 215.

## VI

*Internal/external & double effect.* Quine's attack on internal/external begins with his anti-reductionism, but it doesn't end there. Because up to a point, Carnap *agrees*: any link between theory and observation can be broken, and any can in the right context be forged.<sup>20</sup> It is just that he puts a different spin on these scenarios. There is indeed (thinks Carnap) a possibility that can never be foreclosed. But it is not the possibility of our correcting the rules to accommodate some new finding about the conditions under which *X*-statements are 'really true',<sup>21</sup> it is that we should decide for *practical* reasons to trade the going framework for another, thereby imbuing '*X*' with a new and different meaning.<sup>22</sup>

That Carnap to this extent *shares* Quine's anti-reductionism forces Quine to press his objection from the other side. Having previously argued that the 'internal' life, in which we decide between particular statements, is a looser and more pragmatic affair than Carnap paints it, he needs now to argue that the 'external' life, in which we decide between frameworks, is more evidence-driven and theoretical.

Imagine that the choice before me is whether to adopt a rule making 'there are *X*s' assertible under such and such observational conditions. And assume, as may well be the case, that these conditions are known to obtain; they might obtain trivially, as when '*X*' = 'number'. Then my decision is (in part) a decision about whether to say 'there are *X*s'. Since Carnap gives no hint that these words are to be uttered with anything less than complete sincerity, what I am really deciding is whether to regard 'there are *X*s' as *true* and to *believe* in *X*s.<sup>23</sup> How then does adopting the rule fall short of being the acceptance of new doctrine?

Carnap could play it straight here and insist that adopting the rule involves only a *conditional* undertaking to assent to 'there are *X*s' under specified observational conditions, while adopting the doctrine is categorically aligning myself with the view that there are *X*s. But this is the kind of manoeuvre that gives the doctrine of double effect a bad name. Surely the decision to  $\phi$  cannot disclaim all responsibility for  $\phi$ 's easily foreseeable (perhaps analytically foreseeable) consequences? To portray adopting the rule as taking a stand on what I am going

<sup>20</sup> It is too often forgotten where Quine *gets* his anti-reductionism: 'The dogma of reductionism survives in the supposition that each statement, taken in isolation from its fellows, can admit of confirmation or infirmation at all. My countersuggestion, issuing essentially from Carnap's doctrine of the physical world in the *Aufbau*, is that our statements about the external world face the tribunal of sense experience not individually but only as a corporate body' (Quine 1951a, 41).

<sup>21</sup> There is no scope for such a finding, since there is no external vantage point from which *X*-statements can be evaluated.

<sup>22</sup> This was Carnap's view already in the 1930s: 'all rules are laid down with the reservation that they may be altered as soon as it seems expedient to do so' (Carnap 1937, 318).

<sup>23</sup> 'The acceptance of the thing language leads, on the basis of observations made, also to the acceptance, belief, and assertion of certain statements' (Carnap 1956, 208).

to *mean* by 'X', as opposed to a stand on the facts, is just another version of the same manoeuvre; it is not going to make much of an impression on the man who called it 'nonsense, and the root of much nonsense, to speak of a linguistic component and a factual component in the truth of any individual statement'.<sup>24</sup>

## VII

*Internal/external & pragmatism.* Carnap has his work cut out for him. Can he *without* appeal to analytic/synthetic, and *without* assuming the separability of meaning and 'how things are' as factors in truth, explain why the adoption of new assertion rules is not a shift in doctrine?

He might try the following. *If* the decision to make 'there are Xs' assertible were based in some independent insight into the ontological facts, or even in evidence relevant to those facts, then yes, it would probably deserve to be called a change of doctrine. If anything has been learned, though, from the long centuries of wheel-spinning debate, it is that independent insight and evidence are lacking. The decision to count 'there are Xs' assertible has got to be made on the basis of *practical* considerations: efficiency, simplicity, applicability, fruitfulness, and the like. And what practical considerations rationalize is not change in doctrine, but change in action or policy.

This is where push famously comes to shove. Efficiency and the rest are *not* for Quine 'practical considerations', not if that is meant to imply a lack of evidential relevance. They are exactly the sorts of factors that scientists point to as favouring one theory over another, hence as supporting this or that view of the world. As he puts it in the last sentence of 'Carnap's Views on Ontology', 'ontological questions [for Carnap] are questions not of fact but of choosing a convenient conceptual scheme or framework for science; . . . with this I agree only if the same be conceded for every scientific hypothesis'.<sup>25</sup>

A three-part objection, then: anti-reductionism, double effect, and finally pragmatism. The objection ends as it began, by disparaging not the idea of a Carnapian linguistic framework so much as its bearing on actual practice.<sup>26</sup> The special framework-directed attitudes Carnap points to are, to the extent that we have them at all, attitudes we also take towards our theories. Between acceptance

<sup>24</sup> Quine 1951a, 42. The situation here is more complicated than it may look. Until the framework is adopted, 'there are Xs' has no meaning for me. I am thus faced with a package deal: do I want to mean a certain thing by 'there are Xs', and accept 'there are Xs' with that meaning? Since the meaning is not, pre-adoption, mine, it is questionable whether I can be described, pre-adoption, as considering whether there are Xs, or even considering whether to believe that there are Xs.

<sup>25</sup> Quine 1951b, 72.

<sup>26</sup> Quine on the back of his copy of Carnap 1956: 'When are rules really adopted? Ever? Then what application of your theory to what I am concerned with (language now)?' (Creath 1990, 417).

of a *theory* and acceptance of particular theoretical claims, there is indeed not much of a gap. But it is all the gap that is left between external and internal if Quine is right.

## VIII

*Superficiality of the Quinean critique.* Here is Quine's critique in a nutshell. The factors governing assertion are an inextricable mix of the semantic and the cognitive; any serious question about the assertive use of 'X' has to do both with the word's meaning *and* the X-ish facts. Accordingly Carnap's external stance, in which we confront a purely practical decision about which linguistic rules to employ, and his internal stance, in which we robotically apply these rules to determine existence, are both of them philosophical fantasies.

I want to say that even if all of this is correct, Quine wins on a technicality. His objection doesn't embarrass internal/external as such, only Carnap's way of developing the distinction. To see why, look again at the objection's three stages. The 'anti-reductionist' stage takes issue with Carnap's construal of the framework rules as something like analytic. But analyticity is a red herring. The key point about frameworks for Carnap's purposes is that

- (\*) they provide a context in which we are to say  $--X--$  under these conditions,  $=X=$  under those conditions, and so on, entirely without regard to whether these statements are in a framework-independent sense true.

This is all it takes for there to be an internal/external distinction. And it seems just irrelevant to (\*) whether the rules telling us what to say when are conceived as analytically fixed.

Someone might object that analytical fixity was forced on us by semantic autonomy (by the fact that X has no other meaning than what it gets from the rules), and that semantic autonomy is non-negotiable since it is what licenses (\*)'s insouciance about external truth. Numerical calculation does not answer to external facts about numbers for the same reason that players of tag don't see themselves as answerable to game-independent facts about who is really 'IT'; just as apart from the game there's no such thing as being 'IT', apart from the framework there's no such thing as being 'the sum of seven and five'.

But now wait. If the object is to prevent external claims from 'setting a standard' that internal claims would then be expected to live up to, depriving them of all meaning seems like overkill. A more targeted approach would be to *allow* X-talk its external meaning—allow it to that extent to 'set a standard'—but make clear that internal X-talk is not *bound* by that standard. How to make it clear is the question, and this is where the second or 'double effect' stage comes in.

Must internal utterances have the status of assertions? Carnap's stated goal, remember, is to calm the fears of researchers tempted by Platonic languages; he wants to show that 'using such a language does not imply embracing a Platonic ontology but is perfectly compatible with empiricism and strictly scientific thinking'.<sup>27</sup> If the issue is really one of use and access, then it would seem immaterial whether Carnap's researchers are asserting the sentences they utter or putting them forward in some other and less committal spirit.<sup>28</sup> This takes us to the third or 'pragmatic' stage of Quine's critique.

That frameworks are chosen on practical grounds proves nothing, Quine says, since practical reasons can also be evidential. Of course he's right. But why can't Carnap retort that it was the *other* (the non-evidential) sort of practical reason he had in mind—the other sort of practical reason he took to be at work in these cases? The claim Quine needs is that when it comes to indicative-mood speech behaviour, *no other sort of practical reason is possible*. There is no such thing, in other words, as just putting on a way of talking for the practical advantages it brings, without regard to whether the statements it recommends are in a larger sense true. (If there were, Carnap could take *that* as his model for adopting a framework.)

Does Quine allow for the possibility of ways of talking that are useful without being true, or regarded as true? A few tantalizing passages aside,<sup>29</sup> it seems clear that he not only allows for it, he revels in it. The overall trend of *Word & Object* is that a *great deal* of our day to day talk, and a great deal of the talk even of working scientists, is not to be taken ultimately seriously. This is Quine's famous doctrine of the 'double standard'. Intentional attributions, subjunctive conditionals, and so on are said to have 'no place in an austere canonical notation for science',<sup>30</sup> suitable for 'limning the true and ultimate structure of reality'.<sup>31</sup> Quine does not for a moment suggest these idioms are not useful. He goes out of his way to hail them as indispensable, both to the person in the street and the working scientist.<sup>32</sup> When the physicist (who yields to no one in her determination to

<sup>27</sup> Carnap 1956, 206.

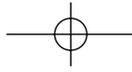
<sup>28</sup> Compare van Fraassen on 'the realist and anti-realist pictures of scientific activity. When a scientist advances a new theory, the realist sees him as asserting the (truth of the) postulates. But the anti-realist sees him as displaying this theory, holding it up to view, as it were, and claiming certain virtues for it' (van Fraassen 1980, 57). A fuller treatment would explore analogies with constructive empiricism; see note 75 for a point of disanalogy.

<sup>29</sup> See especially 'Posits & Reality', originally intended as the opening chapter of Quine 1960. 'Might the molecular doctrine be ever so useful in organizing and extending our knowledge of the behavior of observable things, and yet be factually false? One may question, on closer consideration, whether this is really an intelligible possibility' (Quine 1976, 248). 'Having noted that man has no evidence of the existence of bodies beyond the fact that their assumption helps him organize experience, we should have done well . . . to conclude: such then, at bottom, is what evidence is . . .' (ibid., 251).

<sup>30</sup> Quine 1960, 225.

<sup>31</sup> Quine 1960, 221.

<sup>32</sup> 'Not that I would forswear daily use of intentional idioms, or maintain that they are practically dispensable. But they call, I think, for bifurcation in canonical notation' (Quine 1960, 221).



limn ultimate structure) espouses a doctrine of 'ideal objects' (e.g., point masses and frictionless planes), this is welcomed by Quine as

a deliberate myth, useful for the vividness, beauty, and substantial correctness with which it portrays certain aspects of nature even while, on a literal reading, it falsifies nature in other respects.<sup>33</sup>

Other examples could be mentioned;<sup>34</sup> their collective upshot is that Quine does not really doubt that practical reasons can be given for asserting what are on balance untruths. There is no in-principle mystery (even for him) about the kind of thing Carnap is talking about: a well-disciplined, practically advantageous way of talking that makes no pretence of being 'really true'.

## IX

*What is a framework and what should it be?* About one thing Quine is right. Frameworks cannot remain what they were; they will have to evolve or die. Quine's own view is that he has pushed frameworks in the direction of theories. But his objection really argues, I think, for a different sort of evolution.

Look again at the three stages. The first tells us that frameworks are not to be seen as sole determinants of meaning. All right, let 'X's meaning depend on factors that the framework has no idea of; let 'X' have its meaning quite *independently* of the framework. The second tells us that the rules about what to say when had better not be rules about what to believably assert. All right, let them be rules about what to *put forward*, where this is a conversational move falling short of assertion. The third tells us that if frameworks are non-doctrinal, this is not because they are adopted for reasons like simplicity, fruitfulness, and familiarity. All right, let the conclusion be reached by another and more direct route; let us identify frameworks outright with practices of such and such a type, where it is independently obvious that to engage in these practices is not thereby to accept any particular doctrine.

Now, what is our usual word for an enterprise where sentences are put at the service of something other than their usual truth-conditions, by people who may or may not believe them, in a disciplined but defeasible way? It seems to me that our usual word is 'make-believe game' or 'pretend game'. Make-believe games are

'Not that the idioms thus renounced are supposed to be unneeded in the market place or the laboratory. . . . The doctrine is that all traits of reality worthy of the name can be set down in an idiom of this austere form if in any idiom' (ibid., 228).

<sup>33</sup> Quine 1960, 250.

<sup>34</sup> Just as the immaterialist 'stoop[s] to our [materialist] idiom . . . When the theoretical question is not at issue', and the nominalist 'agree[s] that there are primes between 10 and 20', condoning 'that usage as a mere manner of speaking', many of our own 'casual remarks in the "there are" form would want dusting up when our thoughts turn seriously ontological'. This causes no confusion provided that 'the theoretical use is . . . respected as literal and basic' (Quine 1966a, 99ff).



the paradigm activities in which we ‘assent’ to sentences with little or no regard for their actual truth-values.

Indications are that Carnap would have resisted any likening of the internal to the make-believe. He takes pains to distance himself from those who ‘regard the acceptance of abstract entities as a kind of superstition or myth, populating the world with fictitious . . . entities’.<sup>35</sup> Why, when the make-believe model appears to achieve the freedom from external critique that Carnap says he wants?<sup>36</sup>

First there is a difference of terminology to deal with. A ‘myth’ for Carnap is ‘a false (or dubious) internal statement’—something along the lines of ‘there are ghosts’ conceived as uttered in the thing framework.<sup>37</sup> A ‘myth’ or fiction for me is a *true* internal statement (that is, a statement endorsed by the rules) whose external truth value is as may be, the point being that that truth value is from an internal standpoint quite irrelevant. So while a Carnapian myth *cannot* easily be true, a myth in my sense *must* be internally true and may be externally true as well. (Studied indecision about which of them *are* externally true will be playing an increasing role as we proceed.)

Now, clearly, that ‘internal truths’ are not myths<sub>1</sub> = *statements that pertinent rules of evidence tell us to believe-false* doesn’t show they aren’t myths<sub>2</sub> = *statements that pertinent rules of make-believe tell us to imagine-true*. That said, I suspect that Carnap would not want internal truths to be myths<sub>2</sub> either. This is because freedom from external critique is only part of what Carnap is after, and the negative part at that. There is also the freedom *to* carry on in the familiar sort of unphilosophical way. The internal life Carnap is struggling to defend is the *ordinary* life of the ontologically unconcerned inquirer. And that inquirer does not see herself as playing games, she sees herself as describing reality.

## X

*The effect on Quine’s program.* Playing games vs. describing reality—more on that dilemma in due course.<sup>38</sup> Our immediate concern is not the bearing of

<sup>35</sup> Carnap 1956, 218.

<sup>36</sup> The make-believe interpretation also offers certain advantages. Carnap says that practical decisions as between frameworks are informed by theoretical discussions about ease of use, communicability, and so on. But theoretical statements are always internal, and we are now by hypothesis occupying an external vantage point. Carnap might reply that internal/external is a relative distinction, and that we occupy framework *A* when considering whether to adopt framework *B*. But since the one framework may be just as much in need of evaluation as the other, this makes for a feeling of intellectual vertigo. A cleaner solution is to say that we occupy the external perspective when we in a *non-make-believe* spirit consider the practicality of engaging in make-believe. See also note 47.

<sup>37</sup> Carnap 1956, 218.

<sup>38</sup> I have hopes of enticing the Carnapians back on board by representing it as a false dilemma.

make-believe games on Carnap's program, it's the bearing on Quine's. Quine has not much to say on the topic but it is satisfyingly direct:

One way in which a man may fail to share the ontological commitments of his discourse is . . . by taking an attitude of frivolity. The parent who tells the Cinderella story is no more committed to admitting a fairy godmother and a pumpkin coach into his own ontology than to admitting the story as true.<sup>39</sup>

Note that the imputation of frivolity is not limited just to explicit self-identified pieces of play-acting. Who among us has not slipped occasionally into 'the essentially dramatic idiom of propositional attitudes',<sup>40</sup> or the subjunctive conditional with its dependence on 'a dramatic projection',<sup>41</sup> or the 'deliberate myths'<sup>42</sup> of the infinitesimal and the frictionless plane? Quine's view about all these cases is that we can protect ourselves from ontological scrutiny by keeping the element of drama well in mind, and holding our tongues in moments of high scientific seriousness.

Now, the way Quine is usually read, we are to investigate what exists by reworking our overall theory of the world with whatever tools science and philosophy have to offer, asking all the while what has to exist for the theory to be true. The advice at any particular stage is to

(Q) count a thing as existing iff it is a commitment of your best theory, i.e., the theory's truth requires it.

What though if my best theory contains elements *S* that are there not because they are such very good things to believe but for some other reason, like the advantages that accrue if I *pretend* that *S*? Am I still to make *S*'s commitments my own? One certainly hopes not; I can hardly be expected to take ontological guidance from a statement I don't accept, and may well regard as false!

It begins to look as though (Q) overshoots the mark. At least, I see only two ways of avoiding this result. One is to say that the make-believe elements are never going to make it into our theories in the first place. As theorists we are in the business of describing the world; and to the extent that a statement is something to be pretended true, that statement is not descriptive. A second and likelier thought is that any make-believe elements that do make their way in will eventually drop out. As theory evolves it bids stronger and stronger to be accepted as the honest to God truth. These options are considered in the next few sections; after that we ask what sense can still be made of the Quinean project.

<sup>39</sup> Quine 1961, 103.      <sup>40</sup> Quine 1960, 219.

<sup>41</sup> Dramatic in that 'we feign belief in the antecedent and see how convincing we then find the consequent' (Quine 1960, 222). This hints (quite by accident) at an analogy between the make-believe theory and 'if-thenism' that I hope to pursue elsewhere.

<sup>42</sup> Quine, 248ff.

## XI

*Can make-believe be descriptive?*<sup>43</sup> The thread that links all make-believe games together is that they call upon their participants to pretend or imagine that certain things are the case. These to-be-imagined items make up the game's *content*, and to elaborate and adapt oneself to this content is typically the game's very point.<sup>44</sup>

An alternative point suggests itself, though, when we reflect that all but the most boring games are played with *props*, whose game-independent properties help to determine what it is that players are supposed to imagine. That Sam's pie is too big for the oven doesn't follow from the rules of mud pies alone; you have to throw in the fact that Sam's clump of mud fails to fit into the hollow stump. If readers of 'The Final Problem' are to think of Holmes as living nearer to Hyde Park than Central Park, the facts of nineteenth century geography deserve a large part of the credit.

Now, a game whose content reflects the game-independent properties of worldly props can be seen in two different lights. What ordinarily happens is that we take an interest in the props because and to the extent that they influence the content; one tramps around London in search of 221B Baker Street for the light it may shed on what is true according to the Holmes stories.

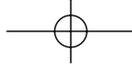
*But in principle it could be the other way around:* we could be interested in a game's content because and to the extent that it yielded information about the props. This would not stop us from playing the game, necessarily, but it would tend to confer a different significance on our moves. Pretending within the game to assert that BLAH would be a way of giving voice to a fact holding *outside* the game: the fact that the props are in such and such a condition, viz., the condition that makes BLAH a proper thing to pretend to assert.

Using games to talk about game-independent reality makes a certain in principle sense, then. Is such a thing ever actually done? A case can be made that it is done all the time—not indeed with explicit self-identified games like 'mud pies' but impromptu everyday games hardly rising to the level of consciousness. Some examples of Kendall Walton's suggest how this could be so:

Where in Italy is the town of Crotona? I ask. You explain that it is on the arch of the Italian boot. 'See that thundercloud over there—the big, angry face near the

<sup>43</sup> This section borrows from Yablo 1997.

<sup>44</sup> Better, such and such is part of the game's content if 'it is to be imagined. . . . *should the question arise*, it being understood that often the question *shouldn't arise*' (Walton 1990, 40). Subject to the usual qualifications, the ideas about make-believe and metaphor in the next few paragraphs are all due to Walton (1990, 1993).



horizon', you say; 'it is headed this way'. . . . We speak of the saddle of a mountain and the shoulder of a highway. . . . All of these cases are linked to make-believe. We think of Italy and the thundercloud as something like pictures. Italy (or a map of Italy) depicts a boot. The cloud is a prop which makes it fictional that there is an angry face . . . The saddle of a mountain is, fictionally, a horse's saddle. But our interest, in these instances, is not in the make-believe itself, and it is not for the sake of games of make-believe that we regard these things as props . . . [The make-believe] is useful for articulating, remembering, and communicating facts about the props—about the geography of Italy, or the identity of the storm cloud . . . or mountain topography. It is by thinking of Italy or the thundercloud . . . as potential if not actual props that I understand where Crotone is, which cloud is the one being talked about.<sup>45</sup>

A certain kind of make-believe game, Walton says, can be 'useful for articulating, remembering, and communicating facts' about aspects of the game-independent world. He might have added that make-believe games can make it easier to reason about such facts, to systematize them, to visualize them, to spot connections with other facts, and to evaluate potential lines of research. That similar virtues have been claimed for metaphors is no accident, if metaphors are themselves moves in world-oriented pretend games:

The metaphorical statement (in its context) implies or suggests or introduces or calls to mind a (possible) game of make-believe . . . In saying what she does, the speaker describes things that are or would be props in the implied game. [To the extent that paraphrase is possible] the paraphrase will specify features of the props by virtue of which it would be fictional in the implied game that the speaker speaks truly, if her utterance is an act of verbal participation in it.<sup>46</sup>

A metaphor on this view is an utterance that represents its objects as being *like so*: the way that they *need* to be to make the utterance pretence-worthy in a game that it itself suggests. The game is played not for its own sake but to make clear which game-independent properties are being attributed. They are the ones that do or would confer legitimacy upon the utterance construed as a move in the game.

Assuming the make-believe theory is on the right track, it will not really do to say that sentences meant only to be pretended-true are nondescriptive and hence unsuited to scientific theorizing. True, to pretend is not itself to describe. But on the one hand, the pretence may only be alluded to, not actually undertaken. And on the other, the reason for the pretence may be to portray the world as holding up its end of the bargain, by being in a condition to make a pretence like that appropriate. All of this may proceed with little conscious attention. Often in

<sup>45</sup> Walton 1993, 40–1.

<sup>46</sup> *Ibid.*, 46. I should say that Walton does *not* take himself to be offering a general theory of metaphor.



fact the metaphorical content is the one that ‘sticks to the mind’ and the literal content takes effort to recover. (Figurative speech is like that; compare the effort of remembering that ‘that wasn’t such a great idea’, taken literally, leaves open that it was a very *good* idea.)

## XII

*Flight from figuration.* What about the second strategy for salvaging (Q)? Our theories may start out partly make-believe (read now metaphorical), but as inquiry progresses the make-believe parts gradually drop out. Any metaphor that is not simply junked—the fate Quine sometimes envisages for intentional psychology—will give way to a paraphrase serving the same useful purposes without the figurative distractions.<sup>47</sup> An example is Weierstrass with his epsilon–delta definition of limit showing how to do away with talk of infinitesimals.

This appears to be the strategy Quine would favour. Not only does he look to science to beat the metaphors back, he thinks it may be the only human enterprise up to the task. He appreciates, of course, that we are accustomed to thinking of ‘linguistic usage as literalistic in its main body and metaphorical in its trimming’. The familiar thought is however

a mistake. . . . Cognitive discourse at its most dryly literal is largely a refinement rather, characteristic of the neatly worked inner stretches of science. It is an open space in the tropical jungle, created by clearing tropes away.<sup>48</sup>

The question is really just whether Quine is *right* about this—not about the prevalence of metaphor outside of science, but about its eventual dispensability within.<sup>49</sup> And here we have to ask what might have drawn us to metaphorical ways of talking in the first place.

A metaphor has in addition to its *literal* content—given by the conditions under which it is true and to that extent belief-worthy—a *metaphorical* content given by the conditions under which it is ‘fictional’ or pretence-worthy in the

<sup>47</sup> The notion of paraphrase has always been caught between an aspiration to symmetry—paraphrases are supposed to *match* their originals along some semantic dimension—and an aspiration to the opposite—paraphrases are supposed to *improve* on their originals by shedding unwanted ontological commitments. (See Alston 1957.) Quine avoids the paradox by sacrificing matching to improvement; he expects nothing like synonymy but just a sentence that ‘serves any purposes of [the original] that seem worth serving’ (Quine 1960, 214). But while this is technically unanswerable, there is still the feeling in many cases that the paraphrase ‘says the same’ as what it paraphrases, or the same as what we were trying to say by its means. A reversion to the poetry-class reading of ‘paraphrase’—a paraphrase of *S* expresses in literal terms what *S* says metaphorically—solves the paradox rather neatly.

<sup>48</sup> Quine 1981, 188–9.

<sup>49</sup> Quine speaks of the ‘inner stretches’ of science; is that to concede that ‘total science’ has no hope of achieving a purely literal state?

relevant game. If we help ourselves to the (itself perhaps metaphorical<sup>50</sup>) device of possible worlds, we can put it like so:

$$S\text{'s } \left\{ \begin{array}{l} \text{literal} \\ \text{metaphorical} \end{array} \right\} \text{ content} =$$

$$\text{the set of worlds that, considered as actual, make } S \left\{ \begin{array}{l} \text{true} \\ \text{fictional} \end{array} \right\}.$$

The role of pretend games on this approach is to warp the usual lines of semantic projection, so as to reshape the region a sentence defines in logical space (Fig. 5.1).<sup>51</sup>

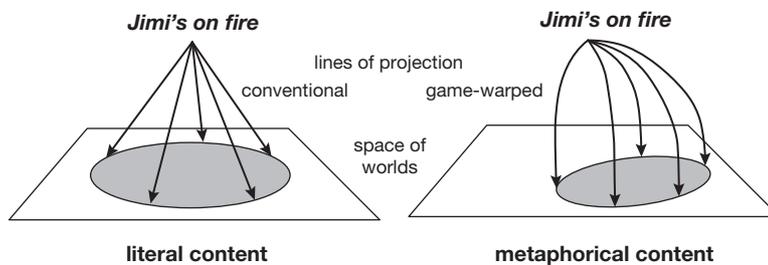


Figure 5.1

The straight lines on the left are projected by the ordinary, conventional meaning of 'Jimi's on fire'; they pick out the worlds which make 'Jimi's on fire' true. The bent lines on the right show what happens when worlds are selected according to whether they make the very same sentence, meaning the very same thing, fictional or pretence-worthy.

If it is granted that there are these metaphorical contents—these ensembles of worlds picked out by their shared property of legitimating a certain pretence—then here is what we want explained: what are the reasons for accessing them *metaphorically*? I can think of at least three sorts of reason, corresponding to three progressively more interesting sorts of metaphor.

### Representationally Essential Metaphors

The most obvious reason is lack of a literal alternative; the language might have no more to offer in the way of a unifying principle for the worlds in a given

<sup>50</sup> Yablo 1997. Derrida was right; one uses metaphor to explain metaphor.

<sup>51</sup> A lot of metaphors are literally impossible: 'I am a rock'. Assuming we want a non-degenerate region on the left, the space of worlds should embrace all 'ways for things to be', not just the 'ways things could have been'. The distinction is from Salmon 1989.

content than that *they* are the ones making the relevant sentence fictional. It seems at least an open question, for example, whether the clouds we call *angry* are the ones that are literally *F*, for any *F* other than ‘such that it would be natural and proper to regard them as angry if one were going to attribute emotions to clouds’. Nor does a literal criterion immediately suggest itself for the pieces of computer code called *viruses*, the markings on a page called *tangled* or *loopy*, the glances called *piercing*, or the topographical features called *basins*, *funnels*, and *brows*.

The topic being ontology, though, let’s try to illustrate with an *existential* metaphor: a metaphor making play with a special sort of object to which the speaker is not committed (not by the metaphorical utterance, anyway) and to which she adverts only for the light it sheds on other matters. An example much beloved of philosophers is *the average so-and-so*.<sup>52</sup> When someone says that

(S) The average star has 2.4 planets,

she is not quite serious; she is pretending to describe an (extraordinary) entity called ‘the average star’ as a way of really talking about what the (ordinary) stars are like on average. Of course, this *particular* metaphor can be paraphrased away, as follows:

(T) The number of planets divided by the number of stars is 2.4,

But the numbers in *T* are from an intuitive perspective just as remote from the cosmologist’s intended subject matter as the average star in *S*. And this ought to make us, or the more nominalistic among us, suspicious. Wasn’t it Quine who stressed the possibility of unacknowledged myth-making in even the most familiar constructions? The nominalist therefore proposes that *T* is metaphorical too; it provides us with access to a content more literally expressed by

(U) There are 12 planets and 5 stars or 24 planets and 10 stars or . . .<sup>53</sup>

And now here is the rub. The rules of English do not allow infinitely long sentences; so the most literal route of access *in English* to the desired content is *T*, and *T* according to the nominalist is a metaphor. It is only by making *as if*

<sup>52</sup> I am indebted to Melia 1995. Following the example of Quine, I will be using ‘metaphor’ in a very broad sense; the term will cover anything exploiting the same basic semantic mechanisms as standard ‘Juliet is the sun’-type metaphors, no matter how banal and unpoetic.

<sup>53</sup> Why not a primitive ‘2.4-times-as-many’ predicate? Because 2.4 is not the only ratio in which quantities can stand; ‘we will never find the time to learn all the infinitely many [*q*-times-as-many] predicates’, with *q* a schematic letter taking rational substituends, much less the *r*-times-as-long predicates, with *r* ranging schematically over the reals (Melia 1995, 228). A fundamental attraction of existential metaphor is its promise of ontology-free semantic productivity. How real the promise is—how much metaphor can do to get us off the ontology/ideology treadmill—strikes me as wide open and very much in need of discussion.

to countenance numbers that one can give expression in English to a fact having nothing to do with numbers, a fact about stars and planets and how they are numerically proportioned.<sup>54</sup>

### Presentationally Essential Metaphors

Whether you buy the example or not, it gives a good indication of what it would be like for a metaphor to be ‘representationally essential’, that is, unparaphrasable at the level of content; we begin to see how the description a speaker wants to offer of his *intended* objects might be inexpressible until *unintended* objects are dragged in as representational aids.

Hooking us up to the right propositional contents, however, is only one of the services that metaphor has to offer. There is also the fact that a metaphor (with any degree of life at all) ‘makes us see one thing as another’;<sup>55</sup> it ‘organizes our view’<sup>56</sup> of its subject matter; it lends a special ‘perspective’ and makes for ‘framing-effects’.<sup>57</sup> Dick Moran has a nice example:

To call someone a tail-wagging lapdog of privilege is not simply to make an assertion of his enthusiastic submissiveness. Even a pat metaphor deserves better than this, and [the] analysis is not essentially improved by tacking on a . . . list of further dog-predicates that may possibly be part of the metaphor’s meaning . . . the comprehension of the metaphor involves *seeing* this person as a lapdog, and . . . experiencing his dogginess.<sup>58</sup>

The point is not essentially about seeing-as, though, and it is not only conventionally ‘picturesque’ metaphors that pack a cognitive punch no literal paraphrase can match. This is clear already from scientific metaphors like *feedback loop*, *underground economy*, and *unit of selection*, but let me illustrate with a continuation of the example started above.

Suppose that I am wrong and ‘the average star has 2.4 planets’ is representationally accidental; the infinite disjunction ‘there are five stars and twelve planets etc.’ turns out to be perfect English. The formulation in terms of the average star is still on the whole hugely to be preferred—for its easier visualizability, yes, but also its greater suggestiveness (‘that makes me wonder how many moons the average planet has’), the way it lends itself to comparison with other data (‘the

<sup>54</sup> Compare Quine on states of affairs: ‘the particular range of possible physiological states, each of which would count as a case of [the cat] wanting to get on that particular roof, is a gerrymandered range of states that could surely not be encapsulated in any manageable anatomical description even if we knew all about cats. . . . Relations to states of affairs, . . . such as wanting and fearing, afford some very special and seemingly indispensable ways of grouping events in the natural world’ (Quine 1966b, 147). Quine sees here an argument for counting states of affairs (construed as sets of worlds!) into his ontology. But the passage reads better as an argument that the *metaphor* of states of affairs allows us access to theoretically important contents unapproachable in any other way.

<sup>55</sup> Davidson 1978.

<sup>56</sup> Max Black in Ortony 1993.

<sup>57</sup> Moran 1989, 108.

<sup>58</sup> Moran 1989, 90.

average planet has nine times as many moons as the average star has planets'), and so on.<sup>59</sup>

Along with its representational content, then, we need to consider a metaphor's *presentational force*. Just as it can make all the difference in the world whether I grasp a proposition under the heading 'my pants are on fire', grasping it as the retroimage of 'Crotona is in the arch of the boot' or 'the average star has 2.4 planets' can be psychologically important too. To think of Crotona's location as the place it would *need* to be to put it in the arch of Italy imagined as a boot, or of the stars and planets as proportioned the way they would need to be for the average star to come out with 2.4 planets, is to be affected in ways going well beyond the proposition expressed. That some of these ways are cognitively advantageous gives us a second reason for accessing contents metaphorically.

### Procedurally Essential Metaphors

A metaphor with only its propositional content to recommend it probably deserves to be considered *dead*; thus 'my watch has a broken hand' and 'planning ahead saves time' and perhaps even 'the number of Democrats is decreasing'. A metaphor (like the Crotona example) valued in addition for its presentational force is *alive*, in one sense of the term, but it is not yet, I think, all that a metaphor can be. This is because we are still thinking of the speaker as someone with a definite *message* to get across. And the insistence on a message settled in advance is apt to seem heavy-handed. 'The central error about metaphor', says Davidson, is to suppose that

associated with [each] metaphor is a cognitive content that its author wishes to convey and that the interpreter must grasp if he is to get the message. This theory is false . . . It should make us suspect the theory that it is so hard to decide, even in the case of the simplest metaphors, exactly what the content is supposed to be.<sup>60</sup>

Whether or not all metaphors are like this, one can certainly agree that a lot are: perhaps because, as Davidson says, their 'interpretation reflects as much on the interpreter as on the originator';<sup>61</sup> perhaps because their interpretation reflects ongoing real-world developments that neither party feels in a position to prejudge. A slight elaboration of the make-believe story brings this third grade of metaphorical involvement under the same conceptual umbrella as the other two:

Someone who utters *S* in a metaphorical vein is recommending the project of (i) looking for games in which *S* is a promising move, and (ii) accepting the propositions that are *S*'s inverse images in those games under the modes of presentation that they provide.

<sup>59</sup> Similarly with Quine's cat example: the gerrymandered anatomical description *even if available* could never do the cognitive work of 'What Tabby wants is that she gets onto the roof'.

<sup>60</sup> Sacks 1978, 44.

<sup>61</sup> Sacks 1978, 29. I hasten to add that Davidson would have no use for even the unsettled sort of metaphorical content about to be proposed.

The overriding principle here is *make the most of it*;<sup>62</sup> construe a metaphorical utterance in terms of the game or games that retromap it onto the most plausible and instructive contents in the most satisfying ways.

Now, should it happen that the speaker has definite ideas about the best game to be playing with *S*, I myself see no objection to saying that she intended to convey a certain metaphorical message—the first grade of metaphorical involvement—perhaps under a certain metaphorical mode of presentation—the second grade.<sup>63</sup> The reason for the third grade of metaphorical involvement is that one can imagine various *other* cases, in which the speaker's sense of the potential metaphorical *truthfulness* of a form of words outruns her sense of the particular truth(s) being expressed. These include the case of the *pregnant* metaphor, which yields up indefinite numbers of contents on continued interrogation;<sup>64</sup> the *prophetic* metaphor, which expresses a single content whose identity, however, takes time to emerge;<sup>65</sup> and, importantly for us, the *patient* metaphor, which hovers unperturbed above competing interpretations, as though waiting to be told where its advantage really lies.<sup>66</sup>

Three grades of metaphorical involvement, then, each with its own distinctive rationale.<sup>67</sup> The Quinean is in effect betting that these rationales are short-term only—that in time we are going to outgrow the theoretical needs to which they speak. I suppose this means that every theoretically important content will find literal expression; every cognitively advantageous mode of presentation will confer its advantages and then slink off; every metaphorical 'pointer' will be replaced by a literal statement of what it was pointing at. If he has an argument for this, though, Quine doesn't tell us what it is. I therefore want to explore the consequences of allowing that like the poor, metaphor will be with us always.

<sup>62</sup> David Hills's phrase, and idea. <sup>63</sup> This of course marks a difference with Davidson.

<sup>64</sup> Thus, each in its own way, 'Juliet is the sun', 'Eternity is a spider in a Russian bathhouse', and 'The state is an organism'.

<sup>65</sup> Examples: An apparition assures Macbeth that 'none of woman born' shall harm him; the phrase's meaning hangs in the air until Macduff, explaining that he was 'from his mother's womb untimely ripped', plunges in the knife. Martin Luther King Jr. told his followers that 'The arc of the moral universe is long, but it bends toward justice'; recent work by Josh Cohen shows that a satisfyingly specific content can be attached to these words. A growing technical literature on verisimilitude testifies to the belief that 'close to the truth' admits of a best interpretation.

<sup>66</sup> 'Patience is the key to content' (Mohammed).

<sup>67</sup> I don't say this list is exhaustive; consider a fourth grade of metaphorical involvement. Sometimes the point is not to advance a game-induced content but to map out the contours of the inducing game, e.g., to launch a game, or consolidate it, or make explicit some consequence of its rules, or extend the game by adjoining new rules. Thus the italicized portions of the following: 'you said he was a Martian, right? well, *Mars is the angry planet*'; '*the average star has a particular size*—it is so many miles in diameter—but *it is not in any particular place*'; 'that's close to right, but *close only counts in horseshoes*'; '*life is a bowl of cherries*, sweet at first but then the pits'. A fair portion of pure mathematics, it seems to me, consists of just such games-keeping.

## XIII

*Can the program be rjiggered?* An obvious and immediate consequence is that the traditional ontological program of believing in the entities to which our best theory is committed stands in need of revision. The reason, again, is that our best theory may well include metaphorical sentences (whose literal contents are) not meant to be believed. Why should we be moved by the fact that *S* as literally understood cannot be true without *Xs*, if the truth of *S* so understood is not something we have an opinion about?

I take it that any workable response to this difficulty is going to need a way of *sequestering* the metaphors as a preparation for some sort of special treatment. Of course, we have no idea as yet what the special treatment would be; some metaphors are representationally essential and so not paraphrasable away. But never mind that for now. Our problem is much more basic.

If metaphors are to be given special treatment, there had better be a way of telling *which statements the metaphors are*. What is it? Quine doesn't tell us, and it may be doubted whether a criterion is possible. For his program to stand a chance, something must be done to fend off the widespread impression that the boundaries of the literal are so unclear that there is no telling, in cases of interest, whether our assertions are to be taken ontologically seriously.

This is not really the place (and I am not the person) to try to bolster the sceptical impression. But if we did want to bolster it, we could do worse than to take our cue from Quine's attack on the analytic/synthetic distinction in 'Two Dogmas'.

One of his criticisms is phenomenological. Quine says he cannot tell whether 'Everything green is extended' is analytic, and he feels this reflects not an incomplete grasp of 'green' or 'extended' but the obscurity of 'analytic'. Suppose we were to ask ourselves in a similar vein whether 'extended' is metaphorical in 'after an extended delay, the game resumed'. Is 'calm' literal in connection with people and metaphorical as applied to bodies of water, or the other way around—or literal in connection with these and metaphorical when applied to historical eras? What about the 'backs' and 'fronts' of animals, houses, pieces of paper, and parades? Questions like these seem unanswerable, and not because one doesn't understand 'calm' and 'front'.

A second criticism Quine makes is that analyticity has never been explained in a way that enables us to decide difficult cases; we lack even a rough criterion of analyticity. All that has been written on the demarcation problem for metaphor notwithstanding, the situation there is no better and almost certainly worse.

A lot of the criteria in circulation are either extensionally incorrect or circular: often both at the same time, like the idea that metaphors (taken at face value) are outrageously false.<sup>68</sup> The criteria that remain tend to reinforce the impression of large-scale indeterminacy. Consider the ‘silly question’ test; because they share with other forms of make believe the feature of settling only so much, metaphors invite outrageously inappropriate questions along the lines of ‘where exactly is the hatchet buried?’ and ‘do you plan to *drop-forge* the uncreated conscience of your race in the smithy of your soul, or use some alternative method?’ But is it silly, or just mind-bogglingly *naive*, to wonder where *the number of planets* might be found, or how much *the way we do things around here* weighs or how it is coloured? It seems to me that it is silly if these phrases are metaphorical, naive if they are literal; and so we are no further ahead.

The heart of Quine’s critique is his vision of what it is to put a sentence forward as (literally) true. As against the reductionist’s claim that the content of a statement is renderable directly in terms of experience, Quine holds that connections with experience are mediated by surrounding theory. This liberalized vision is supposed to cure us of the *expectation* of a sharp divide between the analytic statements, which no experience can threaten, and the synthetic ones, which are empirically refutable as a matter of meaning.

As it happens, though, we have advanced a similarly liberalized vision of what it is to put a sentence forward as metaphorically true. By the time the third level of metaphorical involvement is reached, the speaker may or may not be saying anything cashable at the level of worlds. This is because a statement’s truth-conditions have come to depend on posterity’s judgment as to what game(s) it is best seen as a move in.<sup>69</sup> And it cannot be assumed that this judgment will be absolute and unequivocal: or even that the judgment will be made, or that anyone expects it to be made, or cares about the fact that matters are left forever hanging.

Strange as it may seem, it is this third grade of metaphorical involvement, supposedly at the furthest remove from the literal, that most fundamentally prevents a sharp delineation of the literal.<sup>70</sup> The reason is that *one* of the contents that my utterance may be up for, when I launch *S* into the world in the make-the-most-of-it spirit described above, is its *literal* content. I want to be understood as meaning what I literally *say* if my statement is literally true—count me a player of the ‘null game’, if you like—and meaning whatever my statement projects

<sup>68</sup> ‘Taken at face value’ means ‘taken literally’; and plenty of metaphors are literally true, e.g. ‘no man is an island’. A general discussion of ‘tests for figuration’ can be found in Sadock’s ‘Figurative Speech and Linguistics’ (Ortony 1993).

<sup>69</sup> There are limits, of course; I should say, posterity’s *defensible* judgment.

<sup>70</sup> It prevents a sharp delineation, not of the literal *utterances*, but of the utterances in which speakers are committing themselves to the literal *contents* of the sentences coming out of their mouths. This indeterminacy would remain if, as seems unlikely, a sharp distinction between literal and metaphorical utterances could be drawn.

onto via the right sort of ‘non-null’ game if my statement is literally false. It is thus indeterminate from my point of view whether I am advancing *S*’s literal content or not.<sup>71</sup>

Isn’t this in fact our common condition? When speakers declare that there are three ways something can be done, that the number of *As* = the number of *Bs*, that they have tingles in their legs, that the Earth is widest at the equator, or that Nixon had a stunted superego, they are more sure that *S* is getting at *something* right than that the thing it is getting at is the proposition that *S*, as some literalist might construe it. If numbers exist, then yes, we are content to regard ourselves as having spoken literally. If not, then the claim was that the *As* and *Bs* are equinumerous.<sup>72</sup>

Still, why should it be a bar to ontology that it is indeterminate from my point of view whether I am advancing *S*’s literal content? One can imagine Quine saying: I always told you that ontology was a long-run affair. See how it turns out; if and when the literal interpretation prevails, that will be the moment to count yourself committed to the objects your sentence quantifies over.

Now though we have come full circle—because how the literality issue turns out depends on how the ontological issue turns out. Remember, we are content to regard our numerical quantifiers as literal precisely if, so understood, our numerical statements are true; that is, precisely if there *really are* numbers. Our problem was how to take the latter issue seriously, and it now appears that Quine is giving us no help with this at all. His advice is to countenance numbers iff the *literal* part of our theory quantifies over them; and to count the part of our theory that quantifies over numbers literal iff there turn out to really be numbers.<sup>73</sup>

<sup>71</sup> Indeterminacy is also possible about whether I am advancing a content at all, as opposed to (see note 67 on the fourth grade of metaphorical involvement) articulating the rules of some game relative to which contents are figured, i.e., doing some gameskeeping. An example suggested by David Hills is ‘there are continuum many spatiotemporal positions’, uttered by one undecided as between the substantival and relational theories of spacetime. One might speak here of a fifth grade of metaphorical involvement, which—much as the third grade leaves it open *what* content is being expressed—takes no definite stand on whether the utterance *has* a content.

<sup>72</sup> ‘When it was reported that Hemingway’s plane had been sighted, wrecked, in Africa, the New York *Mirror* ran a headline saying, “Hemingway Lost in Africa”, the word ‘lost’ being used to suggest he was dead. When it turned out he was alive, the *Mirror* left the headline to be taken literally’ (Davidson 1978, 40). I suspect that something like this happens more often than we suppose, with the difference that there is no conscious equivocation and that it is the metaphorical content that we fall back on.

<sup>73</sup> If literal/metaphorical is as murky as all that, how can it serve Carnapian goals to equate external with literal and internal with metaphorical? Two goals need to be distinguished: Carnap’s ‘official’ goal of making quantification over abstract entities nominalistically acceptable in principle; and his more quizzicalistic goal of construing *actual* such quantification in such a way that nominalistic doubts come to appear ingenuous if not downright silly. The one is served by arranging for the quantification to be clearly, convincingly, and invincibly metaphorical; I have said nothing to suggest that a determined metaphor-maker is dragged against her will into the region of indeterminacy. The other is served by construing our actual quantificational practice as metaphorical-iff-necessary, that is, literal-iff-literally-true.

## XIV

*The trouble with 'really'.* The goal of philosophical ontology is to determine what really exists. Leave out the 'really' and there's no philosophy; the ordinary judgment that there exists a city called Chicago stands unopposed. But 'really' is a device for shrugging off pretences, and assessing the remainder of the sentence from a perspective uncontaminated by art. ('That guy's not *really* Nixon, just in the opera'.) And what am I supposed to do with the request to shrug off an attitude that, as far as I can tell, I never held in the first place?

One problem is that I'm not sure what it would *be* to take 'there is a city of Chicago' more literally than I already do.<sup>74</sup> But suppose that this is somehow overcome; I teach myself to focus with laserlike intensity on the truth value of 'there is a city of Chicago, *literally speaking*'. Now my complaint is different: Where are the methods of inquiry supposed to be found that test for the truth of existence-claims thus elaborated? All of our ordinary methods were designed with the unelaborated originals in mind. They can be expected to receive the 'literally speaking' not as a welcome clarification but an obscure and unnecessary twist.

Quine's idea was that our ordinary methods could be 'jumped up' into a test of literal truth by applying them in a sufficiently principled and long-term way. I take it as a given that this is the one idea with any hope of attaching believable truth values to philosophical existence-claims. Sad to say, the more controversial of these claims are equiposed between literal and metaphorical in a way that Quine's method is powerless to address.<sup>75</sup> It is not out of any dislike for the method—on the contrary, it is because I revere it as ontology's

<sup>74</sup> Or to commit myself to taking it more literally than I already may. I have a slightly better idea of what it would be to commit myself to the literal content of 'the number of *As* = the number of *Bs*'. This is why I lay more weight on a second problem; see immediately below.

<sup>75</sup> Which existence-claims am I talking about here? One finds more of an equipose in some cases than others. These are the cases where the automatic presumption in favour of a literal interpretation is offset by one or more of the following hints of possible metaphoricality. *Insubstantiality*: The objects in question have no more to their natures than is entailed by our conception of them, e.g., there is not much more to the numbers than what follows from the 2nd-order Peano Axioms. *Indeterminacy*: It is indeterminate which of them are identical to which, e.g., which sets the real numbers are. *Silliness*: They give rise to 'silly questions' probing areas the make-believe does not address. *Unaboutness*: They turn up in the truth-conditions of sentences that do not intuitively concern them, e.g., 'this argument is valid' is not intuitively about models. *Paraphrasability*: They are oftentimes paraphrasable away with no felt loss of subject matter; 'there are more *Fs* than *Gs*' captures all we meant by 'the number of *Fs* exceeds the number of *Gs*'. *Expressiveness*: They boost the language's power to express facts about less controversial entities, as in the average star example. *Irrelevance*: They are called on to 'explain' phenomena that would not on reflection suffer by their absence; if all the one-one functions were killed off today, there would still be as many left shoes in my closet as right. *Disconnectedness*: Their lack of naturalistic connections threatens to prevent reference relations and epistemic access. I take it that mathematical objects exhibit these features to a higher degree than, say, God, or theoretical entities in physics.

last, best hope—that I conclude that the existence-questions of most interest to philosophers are moot. If they had answers, (Q) would turn them up; it doesn't, so they don't.

## REFERENCES

- W. Alston, 1958. 'Ontological commitment', *Philosophical Studies*, Vol. 9, No. 1, pp. 8–17
- G. Bird, 1995. 'Carnap and Quine: Internal and External Questions', *Erkenntnis*, Vol. 42, pp. 41–46
- R. Carnap, 1936/7. 'Testability and Meaning', *Philosophy of Science*, Vol. 3, 419–471, and Vol. 4, 1–40
- 1937. *The Logical Syntax of Language* (London: Routledge & Kegan Paul)
- 1956. 'Empiricism, Semantics, & Ontology', in his *Meaning & Necessity*, 2nd edition (Chicago: University of Chicago Press)
- 1969. *The Logical Structure of the World & Pseudoproblems in Philosophy* (Berkeley: University of California Press)
- N. Chomsky & I. Scheffler, 1958–9. 'What is said to be', *Proceedings of the Aristotelian Society*, Vol. LIX, pp. 71–82
- R. Creath, ed. 1990. *Dear Carnap, Dear Van* (Los Angeles: UCLA Press)
- D. Davidson, 1978. 'What metaphors mean', in Sacks 1979
- S. Haack, 1976. 'Some preliminaries to ontology', *Journal of Philosophical Logic*, Vol. 5, pp. 457–474
- D. Hills, 1998. 'Aptness and Truth in Metaphorical Utterance', *Philosophical Topics*, 25, pp. 117–154
- C. Hookway, 1988. *Quine* (Stanford: Stanford University Press)
- F. Jackson, 1980. 'Ontological commitment and paraphrase', *Philosophy*, Vol. 55, pp. 303–315
- J. Melia, 1995. 'On what there's not', *Analysis*, Vol. 55, No. 4, pp. 223–229
- R. Moran, 1989. 'Seeing and Believing: Metaphor, Image, and Force', *Critical Inquiry*, Vol. 16, pp. 87–112
- A. Ortony, 1993. *Metaphor and Thought*, second edition (Cambridge: Cambridge University Press)
- W.V. Quine, 1939. 'A logistical approach to the ontological problem', preprint from (the 'destined never to appear') *Journal of Unified Science*; reprinted in Quine 1961
- 1948. 'On what there is', *Review of Metaphysics*, Vol. II, No. 5, reprinted in Quine 1961
- 1951a. 'Two dogmas of empiricism', *Philosophical Review*, Vol. 60, pp. 20–43, reprinted in Quine 1961
- 1951b. 'On Carnap's views on ontology', *Philosophical Studies*, Vol. II, No. 5, pp. 65–72, reprinted in Quine 1976
- 1960. *Word & Object* (Cambridge: MIT Press)
- 1961. *From a Logical Point of View*, 2nd edition (New York, Harper & Row)
- 1964. 'Ontological reduction and the world of numbers', *Journal of Philosophy*, Vol. 61, reprinted with changes in Quine 1976

- W.V. Quine, 1966a and b. 'Existence and Quantification' and 'Propositional Objects', in *Ontological Relativity and Other Essays* (New York: Columbia University Press)
- 1976. *The Ways of Paradox & Other Essays*, revised edition (New York: Columbia University Press)
- 1979. 'A Postscript on Metaphor', in Sacks 1978, reprinted in Quine 1981
- 1981. *Theories and Things* (Cambridge: Harvard University Press)
- G. Ryle, 1954. *Dilemmas* (London: Cambridge University Press)
- S. Sacks, ed., 1978. *On Metaphor* (Chicago: University of Chicago Press)
- N. Salmon, 1989. 'The Logic of What Might Have Been', *Philosophical Review*, Vol. 98, pp. 3–34
- P.A. Schilpp, 1963. *The Philosophy of Rudolf Carnap* (La Salle, Illinois: Open Court)
- L. Stevenson, 1976. 'On what sorts of thing there are', *Mind*, Vol. 85, pp. 503–521
- B. Stroud, 1984. *The Significance of Philosophical Skepticism* (Oxford: Oxford University Press)
- B. van Fraassen, 1980. *The Scientific Image* (Oxford: Clarendon Press)
- K. Walton, 1990. *Mimesis & Make-Believe* (Cambridge, Mass.: Harvard University Press)
- 1993. 'Metaphor and Prop Oriented Make-Believe', *European Journal of Philosophy*, Vol. 1, No. 1. pp. 39–57
- S. Yablo, 1997. 'How in the World?', *Philosophical Topics*, Vol. 24, No.1 pp. 255–286, reprinted in Yablo (2009) pp. 191–221.



# 6

## Apriority and Existence

### 1. A PARADOX

Fifty years ago, something big happened in ontology. W. V. O. Quine convinced everyone who cared that the argument for abstract objects, if there were going to be one, would have to be a posteriori in nature. And it would have to be an a posteriori argument of a particular sort: an *indispensability* argument representing numbers, to use that example, as entities that ‘total science’ cannot do without.<sup>1</sup>

This is not to say that a priori arguments are no longer attempted. They are, for instance by Alvin Plantinga in *The Ontological Argument*, and Crispin Wright in *Frege and the Conception of Numbers as Objects*. These arguments are put forward, however, with a palpable sense of daring, as though a rabbit were about to be pulled out of a hat. Nobody supposes that there are *easy* proofs, from a priori or empirically obvious premises, of the existence of abstracta.<sup>2</sup> (The only easy existence proof we know of in philosophy is Descartes’ *cogito ergo sum*.)

The paradox is that, if we are to go by what philosophers say in other contexts, this bashfulness about what can be shown a priori is quite unnecessary. Abstract objects are a priori deducible from assumptions that nobody would deny.

*Example* (i). As everyone knows, an argument is valid iff every model of its premises is a model of its conclusions. I have never seen empirical evidence offered for this equivalence so I assume the knowledge is a priori. On the other hand, it is *also* (often) known a priori that such and such an argument is invalid.

David Hills, Ken Walton, Mark Crimmins, Ralph Wedgwood, Ned Hall, John Hawthorne, Peter van Inwagen, Stephen Schiffer, David Chalmers, Kent Bach, Laura Shroeter, Sol Feferman, Thomas Hofweber, David Velleman, Peter Railton—thanks for your comments and advice. Related papers were read at Southern Methodist University, University of Colorado, Brandeis University, Harvard University, Brown University, University of Connecticut, Syracuse University, CSLI, Notre Dame University, and Columbia University.

<sup>1</sup> The classic formulation is Hilary Putnam’s: ‘quantification over mathematical entities is indispensable for science . . . , therefore we should accept such quantification; but this commits us to accepting the existence of the mathematical entities in question’ (1971: 57).

<sup>2</sup> A possible exception is Arthur Prior in ‘Entities’, who comments: ‘This is very elementary stuff—I am almost tempted to apply the mystic word “tautological”—and I apologise for so solemnly putting it forward in a learned journal. But I do not think it can be denied that these things need to be said. For there are people who do not agree with them’ (1976: 26).

From these two pieces of a priori knowledge it follows by elementary logic that there exist certain abstract objects, viz. models.

*Example (ii).* It is a priori, I assume, since observational evidence is never given, that there are as many *F*s and *G*s iff there is a one to one function from the *F*s to the *G*s. It is also known, a posteriori this time, that I have as many left shoes as right. From these two pieces of information it again follows by logic that certain abstract objects exist, viz. functions.

## 2. PLATONIC OBJECTS

So far, so bad. But matters can be made even worse. This is because objects that are *not* abstract, or not obviously so, can be similarly ‘deduced’ on the basis of a priori-looking bridge principles. I have in mind principles like ‘it is possible that *B* iff there is a *B*-world’, and ‘Jones buttered the toast *F*-ly iff there was a buttering of the toast by Jones and it was *F*’, and ‘Jones is human iff being human is one of Jones’s properties.’ That non-abstract (or not obviously abstract) objects appear also to admit of overeasy proof shows that we still have not got an exact bead on the problem.

Suppose we try again. There’s a tradition in philosophy of finding ‘unexpected objects’ in truth-conditions—of detecting whatsits in the truth-conditions of statements that are not on the face of it *about* whatsits. So,

<i>the truth-value of</i>	<i>is held to turn on</i>
‘argument <i>A</i> is invalid’	the existence of <i>countermodels</i>
‘it is possible that <i>B</i> ’	the existence of <i>worlds</i>
‘there are as many <i>C</i> s as <i>D</i> s’	the existence of 1–1 <i>functions</i>
‘there are over five <i>E</i> s’	the <i>number</i> of <i>E</i> s exceeding five
‘they did it <i>F</i> ly’	the <i>event</i> of their doing it being <i>F</i>
‘there are <i>G</i> s which BLAH’	there being a <i>set</i> of <i>G</i> s which BLAH
‘she is <i>H</i> ’	her relation to the <i>property H</i> -ness

Objects with a tendency to turn up unexpected in truth-conditions like this can be called *platonic*. Models, worlds, properties, and so on, are platonic, relative to the areas of discourse on the left, because the sentences on the left aren’t intuitively *about* models, worlds, and properties. (If an example of non-platonicness is wanted, consider people in relation to population discourse. That the truth about which regions are populated should hinge on where the people are does not make anything platonic, because people are what population-discourse is visibly and unsurprisingly all about.)

Objects are platonic relative to an area of discourse due to the combination of something positive—the discourse depends for its truth-value on how objects like that behave—with something negative—the discourse is not *about* objects like that. It appears to be this combination, truth-dependence without aboutness,



that makes for the paradoxical result. It appears, in other words, that with *all* platonic objects, abstract or not, there is going to be the possibility of an overeasy existence proof. Just as functions are deducible from my having as many left shoes as right ones, events can be conjured a priori out of the fact that Jones buttered the toast slowly, and worlds out of the fact that she could have done it quickly.

### 3. QUINE'S WAY OR THE HIGHWAY

Our paradox is now shaping up as follows. Let  $X$  be whatever sort of platonic object you like: numbers, properties, worlds, sets, it doesn't matter. Then on the one hand we've got

*Quineanism*: to establish the existence of  $X$ s takes a holistic a posteriori indispensability argument;

while on the other hand we've got

*Rationalism*: the existence of  $X$ s follows by 'truths of reason'—a priori bridge principles—from a priori and/or empirical banalities.

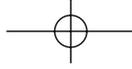
The reason this is a paradox and not merely a disagreement is that Quineanism is received opinion in philosophy, while Rationalism is a straightforward *consequence* of received opinion: the opinion that we are capable in some cases of a priori insight into truth-conditions, and can a priori 'see' that an argument is valid iff it has no countermodels, that it is possible that  $S$  iff there is an  $S$ -world, and so on.

What is to be done? One option of course is to embrace Rationalism and admit that the proof of numbers and the rest is easier than anyone had imagined. I am going to assume without argument that such a course is out of the question. Our feeling of hocus-pocus about the 'easy' proof of numbers (etc.) is really very strong and has got to be respected. If that is right, then only one option remains: we have to renounce our claim to knowing the bridge principles a priori. Perhaps the principles are *false*, as John Etchemendy maintains about the Tarskian validity principles.<sup>3</sup> Or perhaps it is just that our justification is not a priori; the Tarski principle owes its plausibility to the prior hypothesis that there are sets, and the argument for *them* is experiential and holistic. The point either way is that we have to stop carrying on as though it is known independently of experience that, e.g. the valid arguments are the ones without countermodels.

If only it were that easy! The trouble is that our rights of access to the bridge principles do not *seem* to be hostage to empirical fortune in the way suggested;

<sup>3</sup> Etchemendy (1990).





our practice with the principles does not *feel* like it is ‘hanging by a thread’ until the empirical situation sorts itself out. This shows up in a couple of ways, one having to do with our actual attitudes, one having to do with the attitudes we would have in certain counterfactual situations.

*Actual:* Many or most of us using the Tarski biconditional *have no particular view* about abstract ontology. Certainly we are not committed Platonists. If the biconditional (as employed by us) truly presupposed such an ontology, then we *ought* to feel as though we were walking on very thin ice indeed. I don’t know about you, but I have never, not once, heard anxieties expressed on this score.

*Counterfactual:* Also testifying to our (surprising) lack of concern about the true ontological situation is the ‘hypothetical’ fact that if someone were to *turn up* with evidence that abstract objects did not exist, our use of models to figure validity would not be altered one iota. Burgess and Rosen begin their book *A Subject with No Object* with a relevant fable:

Finally, after years of waiting, it is your turn to put a question to the Oracle of Philosophy . . . you humbly approach and ask the question that has been consuming you for as long as you can remember. ‘Tell me, O Oracle, what there is. What sorts of things exist?’ To this the Oracle responds: ‘What? You want the whole list? . . . I will tell you this: everything there is is concrete; nothing there is is abstract.’ (1997: 3)

Suppose we continue the fable a little. Impressed with what the Oracle has told you, you return to civilization to spread the concrete gospel. Your first stop is at—plug in here the name of your favourite department of mathematics or logic—where researchers are confidently reckoning validity by way of calculations on models. You demand that the practice be stopped at once. It’s true that the Oracle has been known to speak in riddles; but there is now a well-enough justified *worry* about the existence of models that all theoretical reliance on them should cease. They of course tell you to bug off and amscray. Which come to think of it is exactly what you yourself would do, if the situation were reversed.

#### 4. IMPATIENCE

Our question really boils down to this. What is the source of the *impatience* we feel with the meddling ontologist—the one who insists that the practice of judging validity by use of Tarski be put on hold until the all-important matter is settled of whether models really exist?

One explanation can be ruled out immediately: we think the principles would still hold (literally) true whether the objects existed or not. That would be to think that if, contrary to what we perhaps suppose, there are no models, then every argument is valid! It would be to think that if the models were found to peter out above a certain finite cardinality—not for deep conceptual reasons, mind you, but as a matter of brute empirical fact—then a whole lot of statements





we now regard as logically contingent, such as ‘there are fifty zillion objects’, are in fact logically false. It seems as clear as anything that we are not in the market for this sort of result. Compare the nonplatonistic bridge principle:

(R) a region is populated iff it contains people.

Should it be discovered that there are no people—everyone but you is a holographic projection, and you are a deluded angel—we would willingly conclude, on the basis of (R), that no regions are populated.<sup>4</sup> And so we can draw the following moral:

*Ontology Matters to Truth:* Our complacency about the bridge principles is *not* due to a belief that they hold literally true regardless of the ontological facts. (It can’t be, since we have no such belief.)

A second explanation of our impatience seems equally misguided: we are confident that the negative empirical findings will never be made. It may be that we *are* confident of this; it is not as though any great number of ontological controversies have been resolved by empirical means in the past. Even if it is granted, though, that we do not expect evidence to turn up that casts doubt on the existence of models, why should that prevent us from having a view about what to say if it did? I take it that we are also confident that it will never be discovered that there are no people. Nevertheless, it seems clearly true that *if* the Oracle convinces us that all the so-called people are no more than clever illusions, we will conclude via the population principle that no region is populated; and clearly false that if the Oracle convinces us that there are no models, we will conclude via Tarski’s principle that all arguments are valid. The point is that

*Experience Matters to Ontology:* Our complacency about the bridge principles is not due to a belief that the trouble-making empirical facts will never come to light. That belief may be there, but our complacency runs deeper than it can explain.

But then it does not really solve the paradox to say that Quineanism wins out over Rationalism. If experience matters to ontology, and ontology matters to truth, then *experience ought to matter to truth* as well. How is it then that the bridge principles are treated, and apparently rightly treated, as experience-independent? What accounts for the a priori-like deference we pay to them? How can we feel justified in *ignoring* a kind of evidence that would, by our own lights, exhibit our belief as false?<sup>5</sup>

<sup>4</sup> This is (one of many reasons) why friends of the population principle do not stay up late worrying about the existence of people.

<sup>5</sup> Here is the problem stated a little more carefully: On the one hand, we feel entitled to the bridge principles regardless of the empirical facts (experience doesn’t matter to truth); on the other hand, we think that the empirical facts are highly relevant to whether the mentioned objects exist (experience does matter to ontology); on the third hand, we think the bridge principles are false if the objects do not exist (ontology matters to truth).



## 5. PLATONISM AS THE PRICE OF ACCESS

Here is the only way out I can see: What entitles us to our indifference about evidence that would exhibit the principles as false is that *we were never committed in the first place to their truth*.<sup>6</sup> Our attitude towards them is attitude *A*, and attitude *A* leaves it open whether the alluded-to objects really exist.

Now that, you may say, is just crazy. Our everyday reliance on the principles surely presupposes a belief in their truth. Take again Tarski's validity principle

(V) an argument is valid iff it has no countermodels.

The point of the 'iff' is to give us licence to reason back and forth between (V)'s left- and right-hand sides, and their negations. If these inferences require us to regard (V) as true, then that is a powerful reason so to regard it.

Humour me for a minute while I state the case a little more guardedly: The back and forth inferences give us reason to regard (V) as true *if* they are inferences that people actually perform.

Well, aren't they? You find a countermodel, you conclude that the argument is invalid. You show that there are no countermodels, you conclude that the argument is valid.

I wonder whether that is a fair description of what really goes on. If you're anything like me, the activity you call 'finding a countermodel' *really* just consists in describing to yourself what the countermodel would have to be like; it consists in laying out a blueprint for a structure of the appropriate sort. The issue of whether anything indeed *answers* to the blueprint is not taken up and seems rather beside the point.

As for the other direction, where countermodels cannot be found and we judge the argument to be valid, again, the activity of 'finding that there are no countermodels' is misdescribed. The fact that one is *really* relying on in judging validity is not that countermodels fail to exist—*that* you could have learned from the Oracle, and it would not have altered your validity-judgements one bit—but that there is something in the very notion of a countermodel to argument *A* that prevents there from being such a thing. A consistent blueprint can't be drawn up because the conditions such a model would have to meet are directly at odds with each other. Once again, the issue of whether models do or do not really exist is not broached and seems of no genuine relevance.

So: if you look at the way the Tarski biconditional is actually used, any larger issue of the existence of models 'in general' is bracketed. It's almost as though we were understanding (V) as

<sup>6</sup> To their literal truth, that is; see below.



(V\*) an argument  $A$  is valid iff—ontological worries to the side, that is, *assuming that models in general exist*— $A$  has no countermodels.<sup>7</sup>

The idea that (V) is in practice understood along the lines of (V\*) has the added virtue of explaining our impatience with the ontologist’s meddling. If the issue is whether there are countermodels *assuming models*, it doesn’t *matter* whether models exist. Of course, the question will be raised of why someone would utter (V) when what they really literally meant was (V\*). Suffice it for now to say that linguistic indirection of this sort is not unknown; we’ll come back to this later. Meanwhile we need to look at some other reasons why a literal interpretation of the bridge principles might seem unavoidable. (Readers in a hurry should go straight to Section 9, or even 10.)

## 6. PLATONISM AS THE KEY TO CLARITY

A great goal of analytic philosophy is to make our ideas clear. Of course, the goal is not often achieved to everyone’s satisfaction, but in a few instances there has been undeniable progress. Everyone will agree, I think, that our notions of limit and of continuity are clearer thanks to Weierstrass’s epsilon-delta story; that our notion of cardinality (especially infinite cardinality) was made clearer by Cantor’s explanation in terms of 1–1 functions; that the notion of inductive definability was clarified by the device of quantifying over all sets meeting appropriate closure conditions; and, to return to our favourite example, that our notion of validity was clarified by the appeal to models. This gives us a second reason for insisting on the reality of platonic objects. If we have to quantify over functions, models, sets, etc. to clarify our ideas, and clarification of ideas is a principal goal of analytic philosophy, how can we be expected to reject such quantification and the ontological commitment it carries?

An example will help us to sort the issue out. Recall the controversy sparked by C. I. Lewis’s work in modal logic. What Lewis did was to distinguish a number of modal systems:  $S_1$ ,  $S_2$ ,  $S_3$ , and so on. These systems, at least the ones that attracted most of the attention, differed in their attitude towards formulae like

- (a)  $\Box P \rightarrow \Box \Box P$ ,
- (b)  $\Diamond \Box P \rightarrow \Box P$ , and
- (c)  $\Diamond P \rightarrow \Box \Diamond P$ .

One response to Lewis’s menu of options was to argue about which of the systems was really ‘correct’. But many philosophers preferred to see disputes

<sup>7</sup> Cf. Field in a critical response to Wright: ‘the conceptual truth is [not ‘the number of  $A$ s = the number of  $B$ s iff there are as many  $A$ s as  $B$ s’ but] rather ‘if numbers exist, then . . .’ (Field 1989: 169).



about which system was best as stemming from subtly different ideas of necessity and possibility. The problem was to identify a *kind* of variation in ideas of necessity that would predict the observed differences in modal intuition.

Then came possible worlds semantics. Acceptance of (a) could now be linked with a transitive conception of relative possibility: a world *w'* that *would* have been possible, had possible world *w* obtained, *is* possible. (Likewise, *mutatis mutandis*, for (b) and (c).) The benefits were and remain substantial: fewer spurious ('merely verbal') disagreements, improved semantical self-understanding, fewer fallacies of equivocation, a clearer picture of why model principles fall into natural packages, and so on.

The platonist now argues as follows. If the clarification that confers these benefits requires us to treat modal operators as (disguised) quantifiers over worlds, then that is how we have to treat them; and that means believing in the worlds.

Isn't there something strange about this line of argument? Clarification is more of a cognitive notion than an ontological one; my goal as a clarifier is to elucidate the content of an idea so that it will be easier to tell apart from other ideas with which it might otherwise get confused. But then, how well I have succeeded ought not to depend on ontological matters *except* to the extent that the content of my idea exhibits a similar dependence.

With some ideas—'externalist' ideas—this condition is perhaps satisfied. There may be no way for me to make my idea of water, or of Hillary Clinton, fully clear without bringing in actual water, or actual Hillary.<sup>8</sup> But my ideas of validity and possibility do not *appear* to be externalist in this way. It is strange then to suppose that actual models and worlds would have to be brought in to make them fully clear.

Where does this leave us? The clarificatory powers of platonic objects are not to be doubted. But they do not depend on the objects' actually being there. I can do just as good a job of elucidating my modal concepts by saying

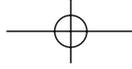
supposing for the moment that necessity is truth at all worlds possible-from-here, *my* concept is one that calls for relative possibility to be transitive,

as I can by saying

*my* concept of necessity has it that necessity *is* truth at all relatively possible worlds, where relative possibility *is* transitive.

Along one dimension, indeed, I can do a better job. Suppose I were to explain my concept of possibility in the second, realistic, way. Then it becomes a conceptual truth that if (contra Lewis) ours is the one and only world, whatever is actually the case is necessarily the case. But this is just *false* of my concept, and I venture

<sup>8</sup> Some would argue that unless there is water, my idea of water cannot *be* fully clear.



to guess of yours as well. An explication that gets a concept's extension-under-a-supposition wrong—that makes mistakes about what goes into the extension on that supposition—does *less* justice to the concept than an explication that avoids the mistakes.

## 7. PLATONISM AS NEEDED FOR PROOF AND EXPLANATION

Another place principle (V) is appealed to is in metalogical proofs. Classical consequence is widely agreed to be monotonic: if  $P_1 \dots P_n/C$  is valid, then so is  $P_1 \dots P_n P_{n+1}/C$ . If we want to prove this result, and/or explain why it holds, we have to quantify over models.<sup>9</sup>

- (i) An argument is valid iff every model of its premises satisfies its conclusion. (This is (V).)
- (ii) If every model of  $P_1 \dots P_n$  satisfies  $C$  then every model of  $P_1 \dots P_n P_{n+1}$  satisfies  $C$ . (By logic and definitions.)
- (iii) If  $P_1 \dots P_n/C$  is valid, then  $P_1 \dots P_n P_{n+1}/C$  is valid. (From (i) and (ii).)

Proofs like this are of course often given. But the reason for giving them is not so clear. It can't be to show *that* monotonicity holds, since on the one hand, no one ever doubted it, while on the other, the Tarskian analysis of validity has been doubted. Nor does the proof do a very good job of explaining *why* monotonicity holds. The fact allegedly being explained—that adding more premises can't make a valid argument invalid—seems on the face of it to lie at a *deeper* level than the facts called in to explain it, that is, the facts stated in (i) and (ii). One might as well try to 'explain' the fact that sisters are siblings by pointing out that a set containing all siblings thereby contains all sisters.

What a proof like the above *does* come close to showing is that monotonicity holds as a conceptual matter, it is implicit in the classical concept of validity.<sup>10</sup> The argument is in two steps. It flows from the classical concept of validity that an argument is valid iff it lacks-countermodels-assuming-models. And it flows from our concept of a model that any countermodel to the 'expanded' argument is a countermodel to the original argument as well. Explicitly:

- (1) An argument is valid iff, assuming models, models of its premises satisfy its conclusion. (This is (V\*), a conceptual truth about validity.)

<sup>9</sup> I am grateful here to Peter van Inwagen, and, for the idea that models are called on to explain validity-facts, to Kent Bach.

<sup>10</sup> As opposed to the various alternative concepts discussed in the literature on nonmonotonic logic.



- (2) Assuming models, if models of  $P_1 \dots P_n$  satisfy  $C$ , then models of  $P_1 \dots P_n P_{n+1}$  satisfy  $C$ . (A conceptual truth about models.)

Now, let it be that  $P_1 \dots P_n / C$  is valid, i.e. that assuming models, models of  $P_1 \dots P_n$  satisfy  $C$ . Then from (2) we see that, again assuming models, models of  $P_1 \dots P_n P_{n+1}$  satisfy  $C$  as well. (The principle used here is that if the members of  $\{A, \text{if } A \text{ then } B\}$  are true-assuming-models, then  $B$  too is true-assuming-models.) So by (1),  $P_1 \dots P_n P_{n+1} / C$  is valid.

- (3)  $P_1 \dots P_n / C$  is valid only if  $P_1 \dots P_n P_{n+1} / C$  is valid. (From (1) and (2).)

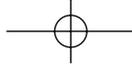
Note that an argument like this is *not* automatically available to someone whose concept of validity is non-classical. Suppose that Smith is working with a version of the ‘circumscriptive’ concept, whereby an argument is valid iff minimal models of its premises are models of its conclusion. Her version of (1)–(3) would start like this:

- (1′) An argument is valid iff, assuming models, *minimal* models of its premises satisfy its conclusion.  
 (2′) Assuming models, minimal models of  $P_1 \dots P_n$  satisfy  $C$  only if minimal models of  $P_1 \dots P_n P_{n+1}$  satisfy  $C$ .

But now wait.  $P_1 \dots P_n$ ’s minimal models may or may not include all of the minimal models of  $P_1 \dots P_n P_{n+1}$ , so (2′) is just false.<sup>11</sup> This illustrates how one can use (1)–(3)-style arguments to tease out the content of a quantificationally explicated concept, without for a moment supposing that the quantified-over entities constitute the real grounds of the concept’s application.

A second example where platonic objects fail to play their advertised role is this. Equinumerosity is symmetrical: if there are exactly as many  $F$ s as  $G$ s, then there are exactly as many  $G$ s as  $F$ s. The usual proof of this result appeals to the fact that inverting a bijection yields another bijection. Do we want to see the proof as *demonstrating*—say, to someone who didn’t already believe it—that exactly-as-many-as is symmetrical? Probably not; that as many  $F$ s as  $G$ s means as many  $G$ s as  $F$ s seems *prima facie* at least as obvious as the invertibility of bijections. Nor does the proof appear to show *why* equinumerosity is symmetrical. If bijections exist, there are going to be lots of them. But then, rather than grounding my fingers’ equinumerosity with my toes in the fact that there are all these bijections, it would seem better to explain the bijections—their possibility, at least—in terms of the prior fact that I have as many fingers as toes. That way we explain many facts in terms of one, rather than one in terms of many.

<sup>11</sup> e.g. let  $P_1 = Fa$ ,  $P_2 = Gb$ , and  $P_3 = \neg Fb$ . Then minimal models of  $\{P_1, P_2, P_3\}$  have two elements each, while those of  $\{P_1, P_2\}$  have just one.



The proof motive for positing platonic objects is not without merit. Platonic argumentation can be enormously instructive.<sup>12</sup> Once we get clearer, though, on what the arguments actually show—not that weakening holds, or that equinumerosity is symmetrical, but that these results are implicit in concepts open to a certain sort of elucidation—then the case for actually *believing* in the objects is tremendously weakened. Once again, we gain as much purchase on the concept by aligning it with a condition on assumed objects as we would by treating the objects as real.

#### 8. PLATONISM AS A CHECK ON PRIMITIVE IDEOLOGY

Everywhere in philosophy we are faced with ‘ideology—ontology’ trade-offs. Roderick Chisholm trades primitive adverbial modification off against sense data; the adverbs win. Donald Davidson trades primitive adverbs off against events; this time the adverbs lose. Arthur Prior has primitive non-nominal quantifiers trading off against properties and propositions. David Lewis pits primitive metaphysical possibility against concrete worlds, conceived as possibility-exemplifiers. Hartry Field does the same, except that his modality is a logical one and the exemplifiers are Tarskian models.

If the examples do nothing else, they remind us that how these trade-offs are carried out is a matter of taste. Some philosophers (e.g. Lewis) want to minimize semantic primitives at the expense of a bigger than expected ontology. Other philosophers (e.g. Field) want to minimize ontology at the expense of a bigger than expected lexicon. About the only thing people seem to agree on is that an *infinite* number of semantic primitives would be too many. Thus Davidson:

When we can regard the meaning of each sentence as a function of a finite number of features of the sentence, we have an insight not only into what there is to be learned [in learning a language]; we also understand how an infinite aptitude can be encompassed by finite accomplishments. For suppose that a language lacks this feature; then no matter how many sentences a would-be speaker learns to produce and understand, there will be others whose meanings are not given by the rules already mastered. It is natural to say that such a language is *unlearnable*. . . . we may state the condition under discussion by saying: a learnable language has a finite number of semantical primitives (1984: 8–9).<sup>13</sup>

The relevance of this to the ontology–ideology issue is that oftentimes the only way of keeping the number of semantic primitives down is to postulate a certain

<sup>12</sup> I should stress that we are not talking about the use of, say, models, to prove results explicitly about models. Our interest is in discourses with respect to which the given objects are platonic.

<sup>13</sup> Davidson sees violations of the learnability requirement in the work of Tarski on quotation marks, Church on sense and denotation, Scheffler on indirect discourse, and Quine on belief attributions.



kind of object. Davidson's showcase example, which he wants to make the basis of a new 'method of truth in metaphysics', has already been mentioned; we have to countenance events, he thinks, to get a tractable semantics for adverbs:

[I]t takes an ontology to make [the device] work: an ontology including people for 'Someone fell down and broke his crown', an ontology of events . . . for 'Jones nicked his cheek in the bathroom on Saturday.' It is mildly ironic that in recent philosophy it has become a popular manoeuvre to try to *avoid* ontological problems by treating certain phrases as adverbial. One such suggestion is that we can abjure sense-data if we render a sentence like 'The mountain appears blue to Smith' as 'The mountain appears bluely to Smith.' Another is that we can do without an ontology of intensional objects by thinking of sentences about propositional attitudes as essentially adverbial: 'Galileo said that the earth moves' would then come out, 'Galileo spoke in-a-that-the-earth-moves-fashion'. There is little chance, I think, that such adverbial clauses can be given a systematic semantical analysis without ontological entanglements (1984: 212–13).

If speakers' competence with adverbs is thought of as grounded (potentially, anyway) in a mechanism that derives '*S* VERBED *Gly*' from a deep structure along the lines of 'there was a VERBing with agent *S* which was *G*', then there will be no need to learn separate inference rules for each action-verb VERB and adverb *G*. Both turn into predicates and so their inferential powers are given by the rules of first-order logic.

The trouble with this as an *ontological* argument is that nowhere in Davidson's account is use made of the fact that the events are actually *there*. At most the conclusion is that we, or pertinent subpersonal systems, are set up to *suppose* they are there. Couldn't the supposition be just that: a supposition? Maybe 'the adverb mechanism' derives '*S* VERBED *Gly*' not from

- (i) 'there was a VERBing with agent *S* which was *G*,' but
- (ii) 'doubts about events aside, there was a VERBing which etc.'

Or maybe it derives '*S* VERBED *Gly*' from (i), but a token of (i) inscribed not in the speaker's 'belief box' but her 'suppose box'. At any rate it is very hard to see how the existence-out-there of real VERBings could lend any help to the speaker trying to acquire a language; whatever it is that events are supposed to contribute to the language-acquisition task would seem to be equally contributed by merely supposed events. This is not to say that there are no events—just that one needs a better reason to believe in them than the help they provide with language-learning.

## 9. PLATONISM AS A PROP FOR REALISM

One more try: why would anyone want (V), or any other bridge principle, to be literally true, so that the platonic objects it quantifies over were really there?

One can think of this as a query about the relations between *ontology*, the study of what is, and *alethiology*, the study of what is the case. A lot of people find it plausible and desirable that what is the case should be controlled as far as possible by what is, and what it is like—that, in Lewis’s phrase,<sup>14</sup> *truth should supervene on being*. This is a view that Lewis himself accepts, in the following form: truth is supervenient on what things there are and which perfectly natural properties they instantiate.<sup>15</sup> Since the properties things instantiate are themselves in a broad sense ‘things’, the view is really that *truth is supervenient on what things there are and their interactions, e.g. which instantiate which*.

Although Lewis maintains supervenience about truth quite generally, it is more common to find it maintained of truth in a particular area of discourse; the usual claim is that truth supervenes on being not *globally* but *locally*. It is very often said that what is wrong, or at least different, about evaluative discourse is that there are no moral/aesthetic *properties* out there to settle the truth-value of evaluative utterances. And it is common to hear anti-realism about *F*-discourse identified with the thesis that there is no such property as *F*ness.<sup>16</sup>

This linking of anti-realism with the lack of an associated property is only one symptom of a broader tendency of thought. When truth in an area of discourse is controlled by the existence and behaviour of objects, that is felt to *boost the discourse’s credentials* as fact-stating or objective. The more truth can be pinned to the way a bunch of objects comport themselves, the more *objective* the discourse appears. Talk about possibility feels more objective if its truth-value is controlled by which possible worlds exist. Talk about what happened yesterday, or what will happen tomorrow, feels more objective if its truth-value is controlled by a still somehow lingering past, or a future out there lying in wait for us.<sup>17</sup> And to return to our original example, talk about validity feels more objective if its truth-value is controlled by the existence or not of countermodels.

Why should objects appear to contribute to objectivity in this way? A little more grandiosely, why should *realism*—which holds that an area of discourse is objective—seem to be bolstered by *platonism*—which points to a special ensemble of objects as determining the distribution of truth values?

Realism à la Dummett says that once you get a sentence’s meaning sufficiently clear and precise, its truth-value is settled. The question is, settled by what? As long as this question is left hanging, there’s room for the anti-realist suspicion that we who employ the sentence are exercising an unwholesome influence.

How is the question to be closed? Well, we’ve got to point to *another* part of reality that *monopolizes* the influence on truth-value, leaving no way that we by our attitudinalizing could be playing a role. This is where platonism comes in. The existence of objects, especially external objects, is the paradigm of an issue

<sup>14</sup> Borrowed from John Bigelow. See Lewis (1992).

<sup>15</sup> Lewis (1992).

<sup>16</sup> This is a particular theme of Paul Boghossian’s paper ‘Status of Content’.

<sup>17</sup> Cf. McDowell on yesterday’s rainstorm.

that's *out of our hands*. Either worlds with flying pigs are there, or they're not. Either tomorrow's sea battle awaits us, or it doesn't. Either the countermodels exist, or they don't.

## 10. A DILEMMA

So—there is a strategy, or tendency of thought, that links *realism* in an area of discourse to *platonism*: belief in a special range of objects whose existence and behaviour settles the question of truth. What are we to make of this strategy? I find it deeply suspicious. The added confidence that the objects are supposed to give us about the objectivity of the discourse strikes me as unearned, or unneeded, or both. To see the problem, look again at what the ontologist is telling us:

You may be right that models aren't needed to settle the truth value of *particular* '*A* has a countermodel' claims. These we can read as short for 'assuming models, *A* has a countermodel.' What you need the models for is the objectivity of the form of speech of which '*A* has a countermodel' is an example. If there really are models, then there's an objective fact of the matter about which arguments have countermodels. Take the models away, and all you've got left is the human practice of developing and swapping around model-descriptions. And this practice, not to say it isn't highly disciplined, doesn't provide as objective a basis for validity-talk as bona fide models would.

The reason I find this suspicious can be put in the form of a dilemma. Logicians speak of 'the space of models', the space that allegedly functions via (V) to make discourse about validity especially objective. Do we have a determinate grasp of this space or not? By a determinate grasp, I mean

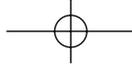
A grasp sufficient to determine a definite truth-value for each instance of 'assuming models, there is a countermodel to argument *A*'.

Does our grasp go fatally blurry, for instance, when it comes to models with very large finite cardinalities? Or is it precise enough to settle the existence of countermodels in every case?

Suppose that it's precise enough; we have a determinate grasp in the specified sense. That by itself ensures that there's a determinate fact of the matter about which arguments have-countermodels-assuming-the-space-of-models.<sup>18</sup>

Suppose next that we *lack* a determinate conception of the space of models; our grasp *fails* to determine an appropriate truth-value for each instance of 'assuming the space of models, there is a countermodel to argument *A*'. How is it that we nevertheless manage to pick out the right class of mathematical objects as models?

<sup>18</sup> Contrast the population principle: region *R* is populated iff there are people in it. A determinate conception of people isn't itself enough to make for an objective fact of the matter about which regions are populated.



The answer has got to be that the world meets us half way. The intended objects somehow jump out and announce themselves, saying: over here, *we're* the ones you must have had in mind. A particularly attractive form of this is as follows: look, we're the only remotely plausible candidates for the job that even *exist*. The idea either way is that we understand the space of models as whatever out there best corresponds to our otherwise indeterminate intentions.

But this reintroduces the hostage-to-fortune problem. An argument's validity-status would seem to be a conceptually necessary fact about it. Surely we don't want the validity of arguments to be held hostage to a brute logical contingency like what model-like entities happen to exist!

So Tarski's principle (considered now as objectively-bolstering) is faced with a dilemma. If we are clear enough about what we *mean* by it, then the principle isn't *needed* for objectivity; (V\*) would do just as well. And if we aren't clear what we mean, then it isn't going to *help*. It isn't even going to be tolerable, because an argument's status as valid is going to blow with the ontological winds in a way that no one could want.

## 11. CRIME OF THE CENTURY?

It begins to look as if the objectivity argument does not really work. The objects would only be needed if they 'stiffened the discourse's spine'—if they had consequences for truth-values over and above anything determined already by our *conception* of the objects. But by that fact alone, we wouldn't trust them to deliver the right results.

The reason this matters is that as far as I can see, the objectivity argument is the *only* one that argues for a truth-link with actual objects. The other principal motives for accepting platonic objects are served just as well by *pretended* or *assumed* ones.

Which suggests a wild idea. Could it be that sets, functions, properties, worlds, and the like, are one and all put-up jobs, meaning, only pretended or assumed to exist? Call this the say-hypothesis, because what it essentially does is construe talk of platonic objects as following on an unspoken 'say there are models (or whatever)' prefix.

How to evaluate the hypothesis? Bertrand Russell complained that postulation of convenient objects has 'all the advantages of theft over honest toil'. This might seem to apply to the say-hypothesis as well. For the suggestion in a way is that an enormous intellectual *crime* has been committed; an entire species of much-beloved and frequently deferred-to entities has been stolen away, leaving behind only persistent appearances.

Suppose we discuss the theft of the platonic objects the way we would any other crime. Means, motive, opportunity—are all these elements present?





The question of means is: how would a job like this be pulled off, where objects appear to be in play but really aren't? The question of motive is: why would anyone *want* to fabricate these objects in the first place? The question of opportunity is: how could a job this big be pulled off without anyone noticing?

## 12. MEANS

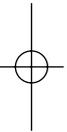
How might it happen that, of the things that regularly crop up in people's *apparently* descriptive utterances, not all really exist, or are even believed to exist by the speaker?

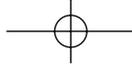
Before addressing this question, we need to acknowledge how nervous it makes us. A certain automatic indignation about people who 'refuse to own up to the commitments of their own speech' has become hugely fashionable. The attitude goes back at least to *Word and Object*, where Quine misses no opportunity to deplore the 'philosophical double talk, which would repudiate an ontology while simultaneously enjoying its benefits' (1960: 242).

But rhetoric aside, the practice of associating oneself with sentences that don't, as literally understood, express one's true meaning is extraordinarily familiar and common. The usual name for it is (not lying or hypocrisy but) but *figurative speech*. I say 'that's not such a great idea' not to call your idea less-than-great—leaving it open, as it were, that it might be very good—but to call your idea bad. The figure in this case is meiosis or understatement. But the point could equally have been made with, say, hyperbole ('they are inseparable'), metonymy ('the White House is angry over allegations that . . .'), or metaphor ('I lost my head'). Not one of the sentences mentioned has a true literal meaning: the first because it exaggerates, the second because it conflates, the third for reasons still to be explored. But it would be insane to associate the speaker with these failings, because the sentences' literal content (if any) is not what the speaker believes, or what she is trying to get across.

The most important example for us is metaphor. What exactly is that? No one quite knows; but the most useful account for our purposes is Kendall Walton's in terms of prop oriented make-believe:

Where in Italy is the town of Crotona? I ask. You explain that it is on the arch of the Italian boot. 'See that thundercloud over there—the big, angry face near the horizon,' you say; 'it is headed this way.' . . . We speak of the saddle of a mountain and the shoulder of a highway . . . All of these cases are linked to make-believe. We think of Italy and the thunder-cloud as something like pictures. Italy . . . depicts a boot. The cloud is a prop which makes it fictional that there is an angry face . . . The saddle of a mountain is, fictionally, a horse's saddle. But . . . it is not for the sake of games of make-believe that we regard these things as props . . . [The make-believe] is useful for articulating, remembering, and communicating facts about the props—about the geography of Italy, or the identity of the storm cloud . . . or mountain topography. It is by thinking of Italy





or the thundercloud . . . as potential if not actual props that I understand where Crotoné is, which cloud is the one being talked about.<sup>19</sup>

A metaphor on this view is an utterance that represents its objects as being *like so*: the way that they would need to be to make it pretence-worthy—or, more neutrally, sayable—in a game that the utterance itself suggests. Sayability here is a function of (a) the rules of the game, and (b) the way of the world. But the two factors play very different roles. The game and its rules are treated as given; they function as medium rather than message. The point of the utterance is to call attention to factor (b), the world. It's to say that *the world has held up its end of the bargain*.

When people talk about metaphor, the examples that come to mind are of metaphorical *descriptions* of everyday objects. A hat is divine; a person is green with envy, or beside herself with excitement. Predicative expressions, though, are far from the only ones we use metaphorically. There is hardly a word in the language—be it an adverb, preposition, conjunction, or what have you—that is devoid of metaphorical potential.

The case of interest to us is *referring phrases*: names, definite descriptions, and quantifiers. An appendix to the *Metaphors Dictionary*<sup>20</sup> lists 450 examples of what it calls 'common metaphors'. Approximately one-half contain referential elements. Some examples drawn just from the beginning of the list:

he fell into *an abyss* of despair, he is tied to *her apron strings*, she has *an axe* to grind, let's put that on *the back burner*, those figures are in *the ballpark*, you're beating *a dead horse*, he's bit off *more than he can chew*, don't hide *your lamp* under *a bushel*, let's go by *the book*, don't blow *a fuse*, I have *a bone* to pick with you, I've burned *my bridges*, I hate to burst *your bubble*, you hit the *bull's-eye*, I have *butterflies* in my stomach, I'm going to lay *my cards* on *the table*, you're building *castles in the air*, we will be under *a cloud* until we settle *this thing*, he claimed *his pound of flesh*, she blew *her cool*, he threw me *a curve*, their work is on *the cutting edge*.

Some additional examples not from the *Dictionary*; with some of them you have to rub your eyes and blink twice before the non-literal aspects shine through:

They put *a lot of hurdles* in your path, there's *a lot* that could be said about that, there's *no precedent* for that, *something* tells me you're right, *there are some things* better left unsaid, *there is something* I forgot to tell you, viz. how to operate the lock, *nothing* gets *my goat* as much as chewing gum in class, *a lot* you can do for me, let's roll out *the red carpet*, *the last thing I want* is to . . . , their people have been rising in *my esteem*, I took her into *my confidence*, *my patience* is nearly exhausted, I'll take *my chances*, there's *a trace of sadness* in your eyes, *a growing number* of these leaks can be traced to Starr's office, she's got *a lot of smarts*, let's pull out *all the stops*; let's proceed along *the lines suggested above*.

Now, the *last thing* I want to do with these examples is to start a bidding war over who can best accommodate our classificatory intuitions. The one

<sup>19</sup> Walton (1993: 40–1).

<sup>20</sup> Sommer and Weiss (1996).





unbreakable rule in the world of metaphor is that there is no consensus on how big that world is: on what should be counted a metaphor and what should not. What I do want to suggest is that the same semantical mechanisms that underlie *paradigmatic* metaphors like ‘your hat is divine’ seem also to be at work with phrases that for whatever reason—too familiar, insufficiently picturesque, too boring—strike us as hardly figurative at all. If that is right, then it does little harm, I think, to *stipulate* that any phrases that turn a non-committal ‘say for argument’s sake that BLAH’ to descriptive advantage are to be seen as just as much metaphorical as the old campaigners.

Pulling these threads together, I contend that the *means* by which platonic objects are simulated is *existential metaphor*—metaphor making play with a special sort of object to which the speaker is not committed (not by the metaphorical utterance, anyway) and to which she adverts only for the light it sheds on other matters. Rather as ‘smarts’ are conjured up as metaphorical carriers of intelligence, ‘numbers’ are conjured up as metaphorical measures of cardinality. More on this below; first there are the questions of motive and opportunity to deal with.

### 13. MOTIVE

What is the *motive* for simulating platonic objects in this way? The answer is that lots of metaphors, and in particular lots of existential metaphors, are *essential*. They have no literal paraphrases: or no readily available ones; or none with equally happy cognitive effects. To see why, we need to elaborate our picture of metaphor a little.

A metaphor has in addition to its literal content—given by the conditions under which it is true and to that extent belief-worthy—a metaphorical content given by the conditions under which it is ‘sayable’ in the relevant game. If we help ourselves (in a purely heuristic spirit)<sup>21</sup> to the device of possible worlds, the claim is that

$$S\text{'s } \left\{ \begin{array}{l} \text{literal} \\ \text{metaphorical} \end{array} \right\} \text{ content} = \text{the set of worlds making } S \left\{ \begin{array}{l} \text{true} \\ \text{sayable} \end{array} \right\}$$

The role of say-games on this approach is to bend the lines of semantic projection, so as to reshape the region a sentence defines in logical space (Fig. 6.1)<sup>22</sup> The straight lines on the left are projected by the ordinary, conventional meaning

<sup>21</sup> Yablo (1996) maintains that worlds are metaphorical. So I am using a metaphor to explain metaphor. Derrida (1982) suggests this is unavoidable. It would be fine by me if he were right.

<sup>22</sup> A lot of metaphors are (literally understood) impossible: ‘I am a rock.’ Assuming we want a non-degenerate region on the left, the space of worlds should take in all ‘ways for things to be’, not just the ‘ways things could have been’. The distinction is from Salmon (1989).



of 'Jimi's on fire'; they pick out the worlds which make 'Jimi's on fire' literally true. The bent lines on the right show what happens when worlds are selected according to whether they make the very same sentence sayable in the relevant game.

The question of motive can now be put like this: granted these metaphorical contents—these ensembles of worlds picked out by their shared property of legitimating an attitude of acceptance-within-the-game—what is the reason for accessing them metaphorically?

One obvious reason would be *lack of an alternative*: the language might have no more to offer in the way of a unifying principle for the worlds in a given content than that *they* are the ones making the relevant sentence sayable. It seems at least an open question, for instance, whether the clouds we call *angry* are the ones that are literally *F*, for any *F* other than 'such that it would be natural and proper to regard them as angry if one were going to attribute emotions to clouds'. Nor does a literal criterion immediately suggest itself for the pieces of computer code called *viruses*, the markings on a page called *tangled* or *loopy*, the vistas called *sweeping*, the glances called *piercing*, or the topographical features called *basins*, *funnels*, and *brows*.

The topic being ontology, though, let's try to illustrate with an *existential*

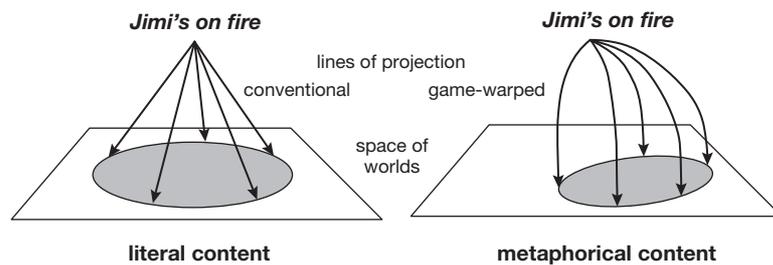


Figure 6.1

metaphor. An example much beloved of philosophers is *the average so-and-so*.<sup>23</sup> When a cosmologist tells us that

(S) The average star has 2.4 planets,

she is not entirely serious; she is making as if to describe an (extraordinary) entity called 'the average star' as a way of really talking about what the (ordinary) stars

<sup>23</sup> I am indebted to Melia (1995). As always I am using 'metaphor' in a very broad sense. The term will cover anything exploiting the same basic semantic mechanisms as standard 'Juliet is the sun'-type metaphors, no matter how banal and unpoetic. (Several people have told me that the semantics of 'average F' is much more complicated than I'm allowing. I am sure they're right, and I apologize for the oversimplification.)

are like on average. True, this *particular* metaphor can be paraphrased away, as follows:

(*T*) The number of planets divided by the number of stars is 2.4.

But the numbers in *T* are from an intuitive perspective just as remote from the cosmologist's intended subject matter as the average star in *S*. And this ought to make us, or the more nominalistic among us, suspicious. Wasn't it Quine who stressed the possibility of unacknowledged myth-making in even the most familiar constructions? The nominalist therefore proposes that *T* is metaphorical too; it provides us with access to a content more literally expressed by

(*U*) There are 12 planets and 5 stars or 24 planets and 10 stars or . . .<sup>24</sup>

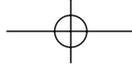
And now here is the rub. The rules of English do not allow infinitely long sentences; so the most literal route of access *in English* to the desired content is *T*, and *T* according to the nominalist is not to be taken literally. It is only by making *as if* to countenance numbers that one can give expression in English to a fact having nothing to do with numbers, a fact about stars and planets and how they are numerically proportioned.<sup>25</sup>

Whether you buy the example or not, it gives a good indication of what it would be like for a metaphor to be 'representationally essential', that is, unparaphrasable at the level of content; we begin to see how the description a speaker wants to offer of his *intended* objects might be inexpressible until *unintended* objects are dragged in as representational aids.

Hooking us up to the right propositional contents, however, is only one of the services that metaphor has to offer. There is also the fact that a metaphor (with any degree of life at all) 'makes us see one thing as another'; it 'organizes

<sup>24</sup> Why not a primitive '2.4-times-as-many' predicate? Because 2.4 is not the only ratio in which quantities can stand; 'we will never find the time to learn all the infinitely many [*q*-times-as-many] predicates', with *q* a schematic letter taking rational substituends, much less the *r*-times-as-long predicates, with *r* ranging schematically over the reals (Melia 1995: 228). A fundamental attraction of existential metaphor is its promise of ontology-free semantic productivity. How real the promise is—how much metaphor can do to get us off the ontology-ideology treadmill—strikes me as wide open and very much in need of discussion.

<sup>25</sup> Compare Quine on states of affairs: 'the particular range of possible physiological states, each of which would count as a case of [the cat] wanting to get on that particular roof, is a gerry-mandered range of states that could surely not be encapsulated in any manageable anatomical description even if we knew all about cats . . . Relations to states of affairs. . . such as wanting and fearing, afford some very special and seemingly indispensable ways of grouping events in the natural world' (Quine 1966: 147). Quine sees here an argument for counting states of affairs into his ontology. But the passage reads better as an argument that the metaphor of states of affairs allows us access to theoretically important contents unapproachable in any other way. See also Lewis on counterfactuals: 'It's the character of our world that makes the counterfactual true—in which case why bring the other worlds into the story at all? . . . it is only by bringing the other worlds into the story that we can say in any concise way what character it takes to make the counterfactual true' (Lewis 1986: 22).



our view' of its subject matter; it lends a special 'perspective' and makes for 'framing-effects'.<sup>26</sup> An example of Dick Moran's:

To call someone a tail-wagging lapdog of privilege is not simply to make an assertion of his enthusiastic submissiveness. Even a pat metaphor deserves better than this . . . the comprehension of the metaphor involves *seeing* this person as a lapdog, and . . . experiencing his dogginess.<sup>27</sup>

The point here is not especially about seeing-as, though, and it is not only conventionally 'picturesque' metaphors that pack the intended sort of cognitive punch. Let me illustrate with a continuation of the example started above.

Suppose I am wrong and 'the average star has 2.4 planets' is representationally *accidental*; the infinite disjunction 'there are five stars and twelve planets etc.' turns out to be perfect English.<sup>28</sup> The formulation in terms of the average star is still on the whole hugely to be preferred—for its easier visualizability, yes, but also its greater suggestiveness ('then how many electrons does the average atom have?'), the way it lends itself to comparison with other data ('2.4 again? Well, what do you know?'), and so on.<sup>29</sup>

A second example has to do with the programme of 'first-orderizing' entailment relations.<sup>30</sup> Davidson in 'The Logical Form of Action Sentences' says that a key reason for rendering 'Jones chewed thoughtfully' as 'there was a chewing done by Jones which was thoughtful' is that the argument from 'Jones chewed thoughtfully' to 'Jones chewed' now becomes quantificationally valid. Of course, similar claims are often made on behalf of the possible worlds account of modality; unless you want the inference from 'possibly *S*' to 'possibly *S-or-T*' to be primitive and unanalyzable, you'd better understand 'possibly *S*' as 'there is a world making *S* true.' Any number of authors have made this sort of plea on behalf of propositions; how without quantifying over them can you hope to first-orderize the inference from 'Tom believes whatever the Pope believes' and 'the Pope believes  $\pi$  is irrational' to 'Tom believes  $\pi$  is irrational'?

The claim these authors make is not that the relevant contents are *inexpressible* without quantifying over events, or worlds, or what have you; that would be untrue, since we can use sentences like 'she VERBED G-ly' and 'possibly BLAH'. It's rather that the logical *relations* among these contents become more tractable if we represent them quantificationally; the contents so represented wear (at least to a first-order-savvy audience like the community of philosophers) their logical potential on their sleeve.<sup>31</sup>

<sup>26</sup> Davidson (1978); Max Black in Ortony (1993); Moran (1989: 108).

<sup>27</sup> Moran (1989: 90).

<sup>28</sup> As maintained, for example, in Langendoen and Postal (1984).

<sup>29</sup> Similarly with Quine's cat example: the gerrymandered anatomical description *even if available* could never do the cognitive work of 'What Tabby wants is that she gets onto the roof.'

<sup>30</sup> See Davidson and Harman (1975). The underlying motivation had to do less with entailment than constructing axiomatic truth theories for natural language. See p. 153.

<sup>31</sup> It is not clear why this presentational advantage should seem to argue for the *truth* of the quantificational rendering, as opposed to just its naturalness and helpfulness. Is it that the



Along with its representational content, then, we need to consider a metaphor's 'presentational force'. Just as it can make all the difference in the world whether I grasp a proposition under the heading 'my pants are on fire', grasping it as the retroimage of 'Crotone is in the arch of the boot' or 'the average star has 2.4 planets' or 'there is a world with blue swans' can be psychologically important too. To think of Crotone's location as the place it would *need* to be to put it in the arch of Italy imagined as a boot, or of the stars and planets as proportioned the way they would need to be for the average star to come out with 2.4 planets, is to be affected in ways going well beyond the proposition expressed. That some of these ways are cognitively advantageous gives us a second reason for accessing contents metaphorically.

#### 14. OPPORTUNITY

Now for the question of opportunity. How are these metaphors slipped in without anyone's noticing?

The first thing that has to be said is that figurative elements in our speech are very *often* unconscious, and resistant to being brought to consciousness. To hear 'that wasn't very smart' (understatement) or 'a fine friend she turned out to be' (irony) or 'spring is just around the corner' (metaphor) as meaning what they literally say takes a surprising amount of effort. A tempting analogy is with the effort involved in making out the intrinsic colour of the paint in some section of a representational painting. As the painting analogy suggests, a too-vivid appreciation of literal meaning can even *interfere* with our understanding of the speaker's message. Wittgenstein imagines an art-lover leaning up to the bloodshot eyes in a Rembrandt painting and saying 'the walls of my room should be painted this colour.' Such a person is not—not at that moment, anyway—in tune with the painting's representational ambitions. Just so, overzealous attention to what a 'gutsy idea' would be like, or what it would really be to 'keep your eyes peeled', or 'pour your heart out' to your beloved, prevents any real appreciation of the intended message.<sup>32</sup>

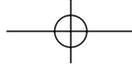
If you're with me this far, consider now statements like 'there's something Jones is that Smith isn't: happy' or 'another way to get there is via Tegucigalpa'? Taken at face value, these sentences do indeed commit themselves to entities called 'happy' and 'via Tegucigalpa'. But overmuch attention to the fact is likelier to distract from the intended meaning than to illuminate it; what on earth could *via Tegucigalpa* be? Likewise someone who says that 'the number of Democrats

naturalness and helpfulness would be a miracle if there were nothing out there answering to the platonic quantifiers? I would like to see an argument for this. I suspect that there are very few putative object-types, however otherwise disreputable, that couldn't be ~~'legitimated'~~ by such a manoeuvre.

<sup>32</sup> Thanks here to Peter Railton.

'legitimated'

^



is on the rise' wants the focus to be on the Democrats, not 'their number', whatever that might be. Their number is called in just to provide a measure of the Democrats' changing cardinality; it's expected to perform that service in the most inconspicuous way and then hustle itself off the stage before people start asking the inevitable awkward questions. (Which number is it? 50 million? Is 50 million really on the rise?)

A deeper reason for the unobtrusiveness of existential metaphors is this. Earlier we distinguished two qualities for which a metaphor might be valued: its representational content, and its presentational force. But that can't be the whole story. For we are still conceiving of the speaker as someone with a definite *message* to get across, and the insistence on a message settled in advance is apt to seem heavy-handed. Davidson says that 'the central error about metaphor' is the idea that

associated with [each] metaphor is a cognitive content that its author wishes to convey and that the interpreter must grasp if he is to get the message. . . . It should make us suspect the theory that it is so hard to decide, even in the case of the simplest metaphors, exactly what the content is supposed to be.<sup>33</sup>

Whether or not all metaphors are like this, one can certainly agree that a lot are: perhaps because, as Davidson says, their 'interpretation reflects as much on the interpreter as on the originator';<sup>34</sup> perhaps because their interpretation reflects ongoing real-world developments that neither party feels in a position to prejudge. Either way, one can easily bring this third, *opportunistic*, grade of metaphorical involvement under the same conceptual umbrella as the other two:

Someone who utters *S* in a metaphorical vein is recommending the project of (i) looking for games in which *S* is a promising move, and (ii) accepting the propositions that are *S*'s inverse images in those games under the modes of presentation that they provide.

The overriding imperative here is to *make the most of it*;<sup>35</sup> we are to construe the utterance in terms of the game or games that retromap it onto the most plausible and instructive contents in the most satisfying ways. Should it happen that the speaker has definite ideas about the best game to be playing with *S*, I myself see no objection to saying that she intended to convey a certain metaphorical message—the first grade of metaphorical involvement—perhaps under a certain metaphorical mode of presentation—the second grade.<sup>36</sup> So it is, usually, with 'He lost his cool (head, nerve, marbles, etc.).'

<sup>33</sup> Davidson (1978: 44).

<sup>34</sup> Ibid. 29. Davidson would have no use for even the unsettled sort of metaphorical content about to be proposed.

<sup>35</sup> David Hill's phrase, and idea.

<sup>36</sup> This of course marks a difference with Davidson.



The reason for the third grade of metaphorical involvement is that one can imagine other cases, in which the speaker's sense of the potential metaphorical truthfulness of a form of words outruns her sense of the particular truth(s) being expressed. Consider, for instance, the *pregnant* metaphor, which yields up indefinite numbers of contents on continued interrogation.<sup>37</sup> Consider the *prophetic* metaphor, which expresses a single content whose identity, however, takes time to emerge.<sup>38</sup> Consider, finally, the *patient* metaphor, which hovers indefinitely above competing interpretations, as though waiting to be told where its advantage really lies.

Strange as it may seem, it is this third grade of metaphorical involvement, supposedly at the furthest remove from the literal, that can be hardest to tell apart from the literal. The reason is that *one* of the contents that my utterance may be up for, when I launch *S* into the world in the opportunistic spirit described above, is its *literal* content. I want to be understood as meaning what I literally say if my statement is literally true (count me a player of the 'null game', if you like) and meaning whatever my statement projects onto via the right sort of 'non-null' game if my statement is literally false. It is thus indeterminate from my point of view whether I am advancing *S*'s literal content or not.<sup>39</sup>

Isn't this in fact our common condition? When people say that the number of apostles is twelve, that rainbows are due to refraction, that Karl Marx had some influential ideas, or that Richard Nixon had a stunted superego, they are far more certain that *S* is getting at *something* right than that the thing it is getting at is the proposition that *S*, as some literalist might construe it. If numbers exist, then yes, we are content to regard ourselves as having spoken literally. If not, then the claim was that there were twelve apostles.<sup>40</sup> If Freud was right, then yes, Nixon had a superego and it really was stunted. If not, then the claim was (more or less) that Nixon had trouble telling when a proposed course of action was morally wrong.

<sup>37</sup> Thus, each in its own way, 'Juliet is the sun' and 'The state is an organism.'

<sup>38</sup> Examples: An apparition assures Macbeth that 'none of woman born' shall harm him; the phrase's meaning hangs in the air until Macduff, explaining that he was 'from his mother's womb untimely ripped', plunges in the knife. Martin Luther King said that 'The arc of the moral universe is long, but it bends towards justice'; Cohen (1997) shows how specific a content can be attached to these words. A growing literature on verisimilitude testifies to the belief that 'close to the truth' admits of a best interpretation albeit one it takes work to find.

<sup>39</sup> Indeterminacy is also possible about whether I am advancing a content at all, as opposed to articulating the rules of some game relative to which contents are figured. An example suggested by David Hills is 'there are continuum many spatio-temporal positions', uttered by one undecided as between the substantival and relational theories of spacetime. One might speak here of a fifth grade of metaphorical involvement, which—much as the third grade leaves it open *what* content is being expressed—takes no definite stand on whether the utterance *has* a content.

<sup>40</sup> 'When it was reported that Hemingway's plane had been sighted, wrecked, in Africa, the New York *Mirror* ran a headline saying, "Hemingway Lost in Africa", the word "lost" being used to suggest he was dead. When it turned out he was alive, the *Mirror* left the headline to be taken literally' (Davidson 1978). I suspect that something like this happens more often than we suppose, with the difference that there is no conscious equivocation and that it is the metaphorical content that we fall back on.



An important special case of the patient metaphor, then, is (what we can call) the *maybe*-metaphor. That platonic metaphors are so often maybe-metaphors—that I *could* for all anyone knows be speaking literally—goes a long way towards explaining their inconspicuousness. If a literal interpretation is always and forever in the offing, then the fact that a metaphorical interpretation is also always and forever possible is liable to escape our notice.

15. . . . LOST?

Of all the reasons people give for thinking that platonic metaphors couldn't have slipped in unnoticed, the most common is probably this. I speak metaphorically only if I speak in a way that is guided by, but somehow at odds with, my notion of what would be involved in a literal deployment of the same sentence.<sup>41</sup> This immediately suggests a negative test. If, as Fowler puts it, metaphors are 'offered and accepted with a consciousness of their nature as substitutes,' then in the absence of any such consciousness—in the absence of a literal meaning the speaker can point to as exploited where it might instead have been expressed—one cannot be speaking metaphorically.

Call this the 'felt distance' test for metaphoricality. It appears to rule that my utterance of, say, 'the number of apostles is twelve' cannot possibly be metaphorical. Were I speaking metaphorically, I would experience myself as guided by meanings of 'number' or 'twelve' that I am at the same time disrespecting or making play with. The fact is, though, that I am not aware of being guided by any such disrespected meanings. I do not even have a conception of what the disrespected meanings could be; it hardly seems possible to use the words 'number' and 'twelve' more literally than I already do.

I have two responses, one which accepts the felt distance test for the sake of argument, one which finds the test unreliable. The first response goes like this. Why do you assume that the words being used metaphorically in 'the number of apostles is twelve' are 'number' and 'twelve'? By a 'number' we mean, roughly: entity of a kind that is suited by its intrinsic nature to providing a measure of cardinality (the number of BLAHs serves as a mark or measure of how many BLAHs there are) and that has not a whole lot more to its intrinsic nature than that. The literal meaning of 'twelve' is: number that provides a measure, cardinality-wise, of the BLAHs just in case there are twelve BLAHs. These are exactly the meanings with which 'number' and 'twelve' are used in 'the number of apostles is twelve'. So it should not be supposed that the metaphoricality of 'the number of apostles is twelve' hinges on a metaphorical usage of those two words.

<sup>41</sup> The intuition here comes out particularly clearly in connection with Walton's account of metaphor; I need first to understand what *S* literally means, if I am to pretend that that meaning obtains in hopes of calling attention to the conditions that legitimate the pretence.



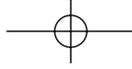
Now, though, the objector will want to know which word *is* being used metaphorically.<sup>42</sup> A plausible candidate is not hard to find. There is a non-negligible chance that numbers do not exist, i.e. that nothing exists whose intrinsic nature is exhausted by the considerations mentioned. Someone who says that ‘the number of apostles is twelve’ is not really concerned about that, however; they are taking numbers for granted in order to call attention to their real subject matter, viz. how many apostles there are. How can someone unconcerned about the existence of *Xs* maintain with full confidence that ‘so and so is the *X* which *Fs*,’ that is, that ‘there is at least one *X* which *Fs* and all such *Xs* are identical to so and so?’ The answer is that they are using the definite article ‘the’, or rather the existential quantifier it implicitly contains, non-literally. Nothing else explains how they can subscribe in full confidence to ‘there is an *X* which *Fs*’ despite being unconvinced of, or at least unconcerned about, the existence of *Xs*. The reason this matters is that the existential quantifier *passes* the felt-distance test. When I assume for metaphorical purposes that numbers exist, I am guided by, but at the same time (running the risk of) disrespecting, the literal meaning of ‘exists’—for using ‘exists’ literally, numbers may well *not* exist, in which case ‘the number of apostles is twelve’, i.e. ‘there is an *x* such that a thing is *x* iff it numbers the apostles and *x* is twelve’, is literally false.

Anyway, though, the felt-distance test is wrong. It is true that if I am to use a sentence *S* metaphorically, there had better be conditions under which *S* is pretence-worthy, or sayable, and conditions under which it is not. But as we know from the example of fiction, this does not require *S* to possess a literal meaning, as opposed to fictionally possessing one in the story or game. Flann O’Brien in *The Third Policeman* tells of a substance called ‘gravid liquid’, the tiniest drop of which weighs many tons, and whose subtle dissemination through the parts of material objects is all that prevents them from floating away. When I pretend, in discussions of that book, that gravid liquid cannot be held in a test tube, since it would break through the bottom, I am guided by my idea of what ‘gravid’ is *supposed in the game* to mean. I have no concern at all about what it means in English, and for all I know it is not even an English word.<sup>43</sup>

^ An example more to the present point is this. ‘Smart’ ~~in my dictionary~~ is an adjective, <sup>or verb</sup> not a noun. How is it that we can say ‘she has a lot of smarts’ and be understood? Well, it is part of the relevant game that there are these entities called ‘smarts’ that are somehow the carriers of intelligence; the more of them you have, the smarter you are. The as-if meaning of ‘smart’ as a noun is of course

<sup>42</sup> I do not see why the weight of a sentence’s metaphoricality should always be borne by particular words. But let’s not get into that here.

<sup>43</sup> Apparently it is; my dictionary gives it the meaning ‘pregnant’. But my use of ‘gravid’ in the game owes nothing to this meaning or any other, or even to ‘gravid’s being a word.



informed by its literal meaning as an adjective. Why should it not be the same with 'twelve'? The meaning it is pretended (or said) to have *qua* noun is informed by its literal meaning *qua* adjective. Much as we're to say that someone has a lot of smarts (noun) just when they're very smart (adjective), we're to say that the number of *F*s is twelve (noun) just when there are twelve (adjective) *F*s.

I don't know which of the two responses to prefer, but let me call attention to a point of agreement between them. A metaphor for us is a supposition adverted to not because it is true but because it marks a place where truths are thought to lie. It is compatible with this that certain words might be used more often in a metaphorical vein than a literal one; it is compatible with it even that certain words should *always* be used metaphorically because they lack literal meaning. This points to a third reason why platonic metaphors do not call attention to themselves.

'Literal' is partly a folk notion, partly a theoretical one. The theoretical idea is that to understand the full range of speech activity, we should employ a divide-and-conquer strategy. Our first step is to set out words' 'primary' powers: what they are in the first instance *supposed* to do. Then we will take on the more multifarious task of accounting for words' 'secondary' powers: their ability to be used in ways not specifically provided for by the primary semantics. A certain kind of Davidsonian, for example, lays great weight on the notion of 'first meaning', constrained by the requirement of slotting into a recursive truth-theory for the full language. Speech is literal if it is produced with intentions lining up in an appropriate way with first meanings; otherwise we have irony, implicature, or metaphor.

Now, to the extent that literality is a theorist's notion, it comes as no great surprise that speakers occasionally misapply it. If we ask the person in the street whether she is using a word literally—using it to do what it is 'supposed' to do—her thoughts are not likely to turn to recursive semantics. More likely she will interpret us as asking about *standard* or *ordinary* usage. (All the more so when an expression has no literal use with which the standard use can be contrasted.) Since platonic metaphors are nothing if not standard, it would be only natural for them to be misconstrued as literal. One doesn't notice that talk of superegos is maybe-metaphorical until one reflects that 'Nixon had a stunted superego' would not be withdrawn even in the proven absence of mental entities with the relevant properties. One doesn't notice that talk of numbers is maybe-metaphorical until one reflects on our (otherwise very peculiar!) insouciance about the existence or not of its apparent objects.

## 16. SUGGESTIVE SIMILARITIES

The bulk of this paper has been an argument that it is less absurd than may initially appear to think that everyday talk of platonic objects is not to be



taken literally. If someone believes that the objects are not really there—that, to revert to the crime analogy, they have been ‘stolen away’—it seems like means, motive, and opportunity for the alleged caper are not at all that hard to make out.

Of course, it is one thing to argue that a metaphorical construal is not out of the question, another to provide evidence that such a construal would actually be correct. The best I can do here is list a series of *similarities* between platonic objects, on the one hand, and creatures of metaphorical make-believe, on the other, that strike me as being, well, suggestive. Not all of the features to be mentioned are new. Not all of them are universal among POs—platonic objects—or MBs—creatures of metaphorical make-believe. Not one of them is so striking as to show decisively that the relevant POs are just MBs. But the cumulative effect is, I think, powerful.

#### PARAPHRASABILITY

MBs are often paraphrasable away with no felt loss of subject matter. ‘That was her first encounter with the green-eyed monster’ goes to ‘that was her first time feeling jealous.’ ‘That really gets my goat’ goes to ‘that really irritates me.’

POs are often paraphrasable away with no felt loss of subject matter. ‘There is a possible world with furry donkeys’ goes to ‘furry donkeys are possible.’ ‘She did it in one way or another’ goes to ‘she did it somehow.’ Etc.

#### IMPATIENCE

One is impatient with the meddling literalist who wants us to get worried about the fact that an MB may not exist. ‘Well, say people *do* store up patience in internal reservoirs; then *my* patience is nearly exhausted.’

One is impatient with the meddling ontologist who wants us to get worried about whether a PO, or type of PO, really exists. ‘Well, say there *are* models; then *this* argument has a countermodel.’

#### TRANSLUCENCY

It’s hard to hear ‘what if there is no green-eyed monster?’ as meaning what it literally says; one ‘sees through’ to the (bizarre) suggestion that no one is ever truly jealous, as opposed, say, to envious.

It is hard to hear ‘what if there are no other possible worlds?’ as meaning what it literally says; one ‘sees through’ to the (bizarre) suggestion that whatever is, is necessary.

INSUBSTANTIALITY

<p>MBs tend to have not much more to them than what flows from our conception of them. The green-eyed monster has no ‘hidden substantial nature’; neither do the real-estate bug, the blue meanies, the chip on my shoulder, etc.</p>	<p>POs often have no more to them than what flows from our conception of them. All the really important facts about the numbers follow from 2nd order Peano’s Axioms. Likewise for sets, functions etc.</p>
---	---

INDETERMINACY

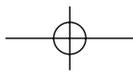
<p>MBs can be ‘indeterminately identical’. There is no fact of the matter as to the identity relations between the fuse I blew last week and the one I blew today, or my keister and my wazoo (‘I’ve had it up to the keister/wazoo with this paperwork’). The relevant game(s) leave it undecided what is to count as identical to what.<sup>44</sup></p>	<p>POs can be ‘indeterminately identical’. There is no fact of the matter as to the identity relations between the pos. integers and the Zermelo numbers, or worlds and maximal consistent sets of propositions, or events and property-instantiations. It is left (partly) undecided what is to count as identical to what.</p>
--	--

SILLINESS

<p>MBs invite ‘silly questions’ probing areas the make-believe does not address, e.g. we know how big the average star is, where is it located? You say you lost your nerve, has it been turned in? Do you plan to <i>drop-forge</i> the uncreated conscience of your race in the smithy of your soul?</p>	<p>POs invite questions that seem similarly silly.<sup>45</sup> What are the intrinsic properties of the empty set? Is the event of the water’s boiling itself hot? Are universals wholly present in each of their instances? Do relations lead a divided existence, parcelled out among their relata?</p>
--	--

<sup>44</sup> ‘Keister’ does in some idiolects have an identifiable anatomical referent; ‘wazoo’ as far as I’ve been able to determine does not. The text addresses itself to idiolects (mine included) in which ‘keister’ shares in ‘wazoo’s’ unspecificity.

<sup>45</sup> Notwithstanding an increasing willingness in recent years to consider them with a straight face. Prior, ‘Entities’, deserves a lot of the credit for this: ‘what we might call Bosanquetterie sprawls over the face of Philosophy like a monstrous tumour, and on the whole the person who maintains that virtue is not square must count himself among the heretics. The current dodge or ‘gambit’ is



## EXPRESSIVENESS

MBs show a heartening tendency to boost the language's power to express facts about other, more ordinary, entities. 'The average taxpayer saves more than the average homeowner.'

POs show a strong tendency to boost the language's power to express facts about other, more ordinary, entities. 'The area of a circle—any circle—is  $\pi$  times the square of its radius.'

## IRRELEVANCE

MBs are called in to 'explain' phenomena that would not on reflection suffer by their absence. 'I curse the HMO because I've had it up to the wazoo with this paperwork.' Take away the wazoos, and people are still going to curse their HMOs.

POs are called in to 'explain' phenomena that would not, on reflection, suffer by their absence. Suppose that all the one-one functions were killed off today; there would still be as many left shoes in my closet as right.

## DISCONNECTEDNESS

MBs have a tendency not to do much *other* than expressive work. As a result, perhaps, of not really existing, they tend not to push things around.

POs show a considerable tendency not to do much other than expressive work. Numbers *et al.* are famous for their causal inertness.

## AVAILABILITY

MBs' lack of naturalistic connections might seem to threaten epistemic access—until we remember that 'their properties' are projected rather than detected.

POs' lack of naturalistic connections might seem to threaten epistemic access—until we recognize that 'their properties' are projected too.

Of course we should not forget one final piece of evidence for the as-if nature of platonic objects. This is the fact that an *as-if interpretation of POs solves our original paradox*. Our reluctance to infer the existence of models from the Tarski equivalences is just what you'd expect if the inference goes through only on a

to say that the question whether virtue is or is not square just doesn't arise, and it is astonishing what a number of questions modern philosophers have been able to dispose of by saying that they just don't arise. Indeed it is hardly too much to say that the whole of traditional philosophy has disappeared in this way, for among questions that don't arise are those which, as it is said, nobody but a philosopher would ask' (1976: 26).



literal interpretation, and Tarski's equation of invalidity with the existence of a countermodel is not in the end taken literally.

## REFERENCES

- Alston, W. (1958), 'Ontological Commitment', *Philosophical Studies* 9(1): 8–17.
- Boghossian, P. (1990), 'The Status of Content', *Philosophical Review* 99(2): 157–84.
- Burgess, J., and G. Rosen (1997), *A Subject With No Object* (Oxford: Clarendon Press).
- Cohen, J. (1997), 'The Arc of the Moral Universe', *Philosophy and Public Affairs* 26(2): 91–134.
- Davidson, D. (1980), *Essays on Actions and Events* (Oxford: Oxford University Press).
- (1984), *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press).
- (1978), 'What Metaphors Mean', in Sacks 1979.
- and G. Harman (1975), *The Logic of Grammar* (Encino: Dickenson).
- Davies, M. and L. Humberstone (1980), 'Two Notions of Necessity', *Philosophical Studies* 38: 1–30.
- Derrida, J. (1982), 'White Mythology: Metaphor in the Text of Philosophy', in *Margins of Philosophy* (Chicago: University of Chicago Press).
- Etchemendy, J. (1990), *The Concept of Logical Consequence* (Cambridge, Mass.: Harvard University Press).
- Field, H. (1989), 'Platonism for Cheap? Crispin Wright on Frege's Context Principle', in *Realism, Mathematics and Modality* (Oxford: Blackwell).
- Hahn, L. and P. Schilpp (eds.) (1986), *The Philosophy of W. V. Quine* (La Salle, Ill.: Open Court).
- Hills, D. (1998), 'Aptness and Truth in Metaphorical Utterance', *Philosophical Topics* 25: 117–154.
- Kaplan, D. (1989), 'Demonstratives', in J. Almog, J. Perry, and H. Wettstein (ed.), *Themes from Kaplan* (New York: Oxford University Press).
- Langendoen, D. and P. Postal (1984), *The Vastness of Natural Languages* (Oxford: Blackwell).
- Lewis, D. (1986), *On the Plurality of Worlds* (New York: Blackwell).
- (1992), 'Critical Notice of D. M. Armstrong, *A Combinatorial Theory of Possibility*', *Australasian Journal of Philosophy* 70: 211–24.
- Maddy, P. (1997), *Naturalism in Mathematics* (Oxford: Clarendon Press).
- McGee, V. (1997), 'How We Learn Mathematical Language', *Philosophical Review* 106: 35–68.
- Melia, J. (1995), 'On what there's not', *Analysis* 55(4): 223–9.
- Moran, R. (1989), 'Seeing and Believing: Metaphor, Image, and Force', *Critical Inquiry* 16: 87–112.
- Ortony, A. (1993), *Metaphor and Thought*, 2nd edn. (Cambridge: Cambridge University Press).
- Plantinga, A. (ed.) (1965), *The Ontological Argument* (Garden City, NY: Doubleday).
- Prior, A. (1976), *Papers in Logic and Ethics* (Amherst: University of Massachusetts Press).
- Putnam, H. (1971), *Philosophy of Logic* (New York: Harper & Row).

- Quine, W. V. (1948). 'On What There is', *Review of Metaphysics* 2(5), repr. in Quine 1953.
- (1953), *From a Logical Point of View* (Cambridge, Mass.: Harvard University Press).
- (1960), *Word and Object* (Cambridge, Mass.: MIT Press).
- (1966), 'Propositional Objects', in *Ontological Relativity and Other Essays* (New York: Columbia University Press).
- (1978), 'A Postscript on Metaphor', in Sacks (1978).
- (1986), 'Reply to Parsons', in Hahn and Schilpp (1986).
- Sacks, S. (ed.) (1978), *On Metaphor* (Chicago: University of Chicago Press).
- Salmon, N. (1989), 'The Logic of What Might Have Been', *Philosophical Review* 98: 3–34.
- Sommer, E. and D. Weiss (1996), *Metaphors Dictionary* (New York: Visible Ink Press).
- Walton, K. (1993), 'Metaphor and Prop Oriented Make-Believe', *European Journal of Philosophy* 1: 39–57.
- Wright, C. (1983), *Frege's Conception of Numbers as Objects* (Aberdeen: Aberdeen University Press).
- Yablo, S. (1996), 'How in the World?' *Philosophical Topics* 24: 255–86.
- (1998), 'Does Ontology Rest on a Mistake?' *Proceedings of the Aristotelian Society* supp. Vol. 72: 229–62. [Chapter 5 in this volume].

## Go Figure: A Path through Fictionalism

### 1. INTRODUCTION

There is the following predicament. One, we find ourselves uttering sentences that seem on the face of it to be committed to so-and-so's—sentences that could not be true unless so-and-so's existed. But, two, we do not *believe* that so-and-so's exist.<sup>1</sup>

What is someone caught up in The Predicament (as let's call it) supposed to do? The official standard menu of options was given by Quine in *Word and Object*. Our choices are three:

- (1) Show how the commitment can be paraphrased away—thus Quine himself on *chances*.
- (2) Stop uttering the problematic sentences—thus Quine on *glints*.
- (3) Acknowledge the commitment—thus Quine on *sets*.

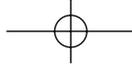
Those who reject these options are subjected by Quine to some pretty withering criticism: “I deplore the philosophical double talk, which would repudiate an ontology while simultaneously enjoying its benefits” (242).

### 2. FOURTH WAY

Quine's menu and the associated moralizing have been terrifically influential. But they have occasioned a fair amount of resentment as well. How do we know the menu is complete? Might there perhaps be some other way of sticking with sentences whose commitments one does not share? Quine of all people should hope so, because he sticks with some himself:

This paper was written for an APA Symposium on Semantic Pretense organized by Mark Richard. Mark Crimmins spoke as well, and the commentators were Thomas Hofweber and Jason Stanley. Thanks, you guys. And thanks to the following for criticism and advice: Gideon Rosen, David Hills, Ken Walton, Bob Stalnaker, Penelope Maddy, Terry Horgan, Tamar Gendler, Peter Ludlow, and Ruth Millikan.

<sup>1</sup> Better, we don't think the propriety of our stance depends on the belief, even if we have it. You may have horses for all I know. But I am not committing myself on the topic when I say you should hold your horses.



I would [not] undertake to limit my use of the words ‘attribute’ and ‘relation’ to contexts that are excused by the possibility of . . . paraphrase . . . consider how I have persisted in my vernacular use of ‘meaning,’ ‘idea,’ and the like, long after casting doubt on their supposed objects. True, the use of a term can sometimes be reconciled with rejection of its objects; but I go on using the terms without even sketching any such reconciliation.<sup>2</sup>

His excuse is that he does not go in for this sort of talk when speaking in “full scientific seriousness,” “limning the true and ultimate structure of reality.” But, how can it excuse an activity to say that one does not go in for it all the time? Quine does speak this way most of the time. And so we are entitled to ask: how do *you* get away with uttering sentences committed to so-and-so’s, where the commitment is not paraphrasable away?

Quine does not say a whole lot about this, and the things he does say do not always fit together. He has on the one hand his doctrine of “the double standard.” Talk about attributes and meanings is excused by its limited ambitions.

What is involved here is simply a grading of austerity. I can object to using a certain dubious term at crucial points in a theory. . . . but I can still use and condone the term in more casual or heuristic connections, where less profundity of theoretical explanation is professed.<sup>3</sup>

If the question is how Quine escapes commitment, the answer is that he does not escape it. He *is* in everyday contexts overcommitted; that’s all right, because he never claimed in these contexts to be limning ultimate structure. This is the “apologist” strand in Quine.

But there has got to be more to the story than that. Statements that do not “limn ultimate structure” are false. And Quine does not think he found us false; he thinks he found us muddled. Ordinary speech is unclear. Our commitments cannot be read off of what we say. Canonical notation is introduced to bring these two back into line. A speaker’s commitments in uttering non-canonical sentences are those of the canonical sentences she is content to put in their place. This is the “hermeneutic” strand in Quine, which seems clearly dominant.

Apologetic Quine thinks talk of glints is false, albeit excusably so. Hermeneutic Quine thinks it may not be false; so far though we have no explanation of how that could be. If I say “there is an X in the closet” when in reality there are no X’s, haven’t I misstated the facts? Maybe, but on the other hand maybe not. It depends on the spirit in which the sentence is put forward.

One way in which a man may fail to share the ontological commitments of his discourse is . . . by taking an attitude of frivolity. The parent who tells the Cinderella story is no more committed to admitting a fairy godmother and a pumpkin coach into his own ontology than to admitting the story as true.<sup>4</sup>

<sup>2</sup> Quine 1960, 210.

<sup>3</sup> Quine 1960, 210.

<sup>4</sup> Quine 1961, 103.



Is it possible this is meant to apply more broadly, to statements that are not so obviously phony?

Once again we get only hints. The language of belief attribution is for Quine an “essentially dramatic idiom.”<sup>5</sup> The subjunctive conditional depends on “a dramatic projection,” in that we are called on to “feign belief in the antecedent.”<sup>6</sup> He speaks of the “deliberate myths”<sup>7</sup> of the infinitesimal and the frictionless plane. Quine’s view about these cases resembles his view of fairy tales. He thinks that we can protect ourselves from ontological scrutiny by keeping the element of drama well in mind and holding our tongues when the mood turns scientific. It appears then that Quine recognizes a *fourth* way of dealing with The Predicament. Someone whose sentences are committed to so-and-so’s need not share in the commitment if

(4) the sentences are advanced in a fictional or make-believe spirit.

To have a name for this fourth option, let us call it *fictionalism*. There are a number of versions of fictionalism, according to the various accounts one might give of “advancing in a fictional spirit.”

### 3. INSTRUMENTALIST FICTIONALISM

The fictionalist holds that we “make as if” we are asserting that S and/or believing that S and/or receiving the news that S. Our reason for making as if we are doing these things (assuming we have a reason—more on this below) is that it serves some larger purpose. Making as if S enables us to simplify our theory, or shorten proofs.

Someone who stops here—someone with no story to tell about what we are “really” doing in making as if S, and why that would be a sensible thing to do—I will call an *instrumentalist fictionalist*, or simply an *instrumentalist*.

I see three main problems for instrumentalism. The first is phenomenological. When I say that “there is a world in which donkeys fly” or “the number of apostles is twelve” or “ $2 + 3 = 5$ ,” these utterances *seem* to mark genuine beliefs of mine, beliefs that I am trying to express and, if possible, communicate to others. If I am not sincerely asserting that the number of apostles is identical to the number twelve (I do not believe in numbers), I do seem to be sincerely asserting something. What is it? The instrumentalist doesn’t say. Call this the problem of *real content*.

A second and related problem is that my utterances would seem to be characterizable as correct or incorrect. It is correct to say that  $2 + 3 = 5$ ,

<sup>5</sup> Quine 1960, 219.

<sup>6</sup> Full quotation: “We feign belief in the antecedent and see how convincing we then find the consequent” (Quine 1960, 222).

<sup>7</sup> Quine 1960, 248ff.

incorrect to say that that  $2 + 3 = 6$ . The problems are connected in that, intuitively anyway, my utterance is correct iff its real content is true. By ignoring the real content side of the equation, the instrumentalist would seem to leave herself with no good way of distinguishing correct utterances from incorrect. Call this the problem of *correctness*.

The third problem is pragmatism. How is the instrumentalist going to fight off the Quine/Putnam objection that says: It quacks like a duck, so it is a duck. You *say* you do not really believe these sentences. But that certainly is not the way you act. You put the sentences out there, you get defensive when people question them, you engage in evidence-gathering and proof-checking and all the rest. At the very least you owe us an account of how all this falls short of belief.

The account one would like to give is as follows. It is true that I carry on like a believer. But, then, I am a believer. I believe the real contents of the sentences you hear me uttering. If people “deny” that the number of apostles is twelve, they are speaking not to the existence of numbers but to how many apostles there are. Naturally then I get mad! They are denying something that I think is really the case, viz., that there are twelve apostles.

An example of an instrumentalist fictionalist is Hartry Field in *Science without Numbers*. Field has us quasi-asserting various things about mathematical objects because to do so shortens our proofs of claims about regular concrete objects. He does not, as far as I know, say that to quasi-assert  $S$  is to really assert something else  $S^*$ .

An example of a *non-instrumentalist* fictionalist is Bas van Fraassen in *The Scientific Image*. Van Fraassen says that we quasi-assert various things about unobservables because this improves our ability to organize and derive results about observables. If that were the whole story, then van Fraassen would be an instrumentalist. But he says more: “When a scientist advances a new theory, the realist sees him as asserting the (truth of the) postulates. But the anti-realist sees him as displaying this theory, holding it up to view, as it were, *and claiming certain virtues for it*” (van Fraassen 1980, 57). One quasi-asserts the theory, but *really asserts* that it has certain virtues, such as empirical adequacy.

#### 4. META-FICTIONALISM

Fictionalists usually have a story to tell about how correct statements differ from incorrect ones. Van Fraassen thinks a scientific statement is correct iff it is part of a theory with such and such virtues, among them empirical adequacy. Field thinks that a mathematical statement is correct iff it follows from standard mathematics. Schematically, we can say that  $S$  is correct iff  $C(S)$ , where  $C$  is the condition the fictionalist puts forward as making for correctness. The difference between van Fraassen and (early) Field is that van Fraassen sees a quasi-assertion that  $S$  as at the same time a genuine assertion that  $C(S)$ . (Field changes his view



about this in the Introduction to *Realism, Math, and Modality*. Quasi-asserting that  $2 + 3 = 5$  is, or can be, really asserting that according to standard math,  $2 + 3 = 5$ .)

So now we have a second sort of fictionalism, favored by van Fraassen and (later) Field. It says that in making as if to assert that S, one is really asserting that S is the right kind of thing to make as if to assert: the quasi-assertion game one is involved in *endorses* the quasi-assertion that S. A natural label would be *meta-fictionalism*, for the real content concerns a sentence and its property of being a good or approved thing to say.

One problem with meta-fictionalism is modal in character. It ordinarily strikes us that  $2 + 3$  is necessarily 5; it could not have been otherwise. But it could (perhaps) have been otherwise that  $2 + 3 = 5$  according to standard math. For standard math could (perhaps) have been different. Certainly it is not *a priori* that standard math turned out the way it did. It is not *a priori* then that according to standard math, there are everywhere continuous functions that are nowhere differentiable. That there are functions like that is, however, the kind of thing mathematicians take themselves to know *a priori*.

Second, there are problems of concern, of what we care about. It is a matter of concern that the number of starving people is large and rising. We do not seem to care in the same way about the content of standard math. If the meta-fictionalist is correct, though, quasi-asserting that the number of starving people is large is really asserting that the number is large *according to standard math*. And it does not seem plausible that what we regret or deplore or are concerned about here is that the number is large according to standard math. Because that would be to deplore *inter alia* part of the content of standard math.

Third is a phenomenological worry. When we say that the number of starving people is very large, we do not feel ourselves to be talking (even a bit!) about the content of a mathematical story. Our subject matter is *people* and our thought is that a lot are starving. “The number of apostles is twelve” is no more about a story than “snow is white” concerns the rules of English.

## 5. OBJECT FICTIONALISM

I am certainly *relying* on the rules of English when I utter the words “snow is white.” It is those rules that make my utterance a way of saying that snow is white. It is just that relying on rules is one thing; talking about those rules is another. Likewise, when the words “the number of apostles is twelve” come out of my mouth, I am relying on the number-fiction. It is thanks in part to that fiction that my utterance is a way of saying that there are twelve apostles. Again, though, relying on a fiction is one thing; talking about it is another. The fiction (like the rules of English) functions as medium and not message.



Now, the rules of English make their contribution roughly like so. The rules tell us which sentences are true under which worldly conditions. If  $K$  is a condition sufficient by  $R$ 's lights for the truth of  $S$ , we write  $R^K \geq S$ . If  $K$  is necessary by  $R$ 's lights for the truth of  $S$ , we write  $R^K \leq S$ .  $R^K = S$  then means that  $K$  is exactly what is needed for  $S$  to come out true, where truth is judged according to  $R$ . The condition that  $R$  sets for the truth of  $S$  is the (literal) content of  $S$ . So

$$(A) \text{ litcontent}(S) = [\text{the } K \text{ such that } R^K = S]$$

There is terminology for this in mathematics. The exponent  $k$  to which you have to raise  $r$  to obtain  $s$  is  $\log_r(s)$  = the logarithm of  $s$  to the base  $r$ . So another way to put it is

$$(B) \text{ litcontent}(S) = \log_R(S)$$

Note the contribution of  $R$ . The rules associate a particular content with  $S$ , but they do not figure in that content. (B) does not implicate  $R$  in the content any more than  $2 = \log_3 9$  implicates 3 in 2.

The role just envisaged for  $R$  is the role we want fiction  $F$  to play as well. The real content of  $S$  is not that  $S$  is fictional, that is, true according to  $F$ . The real content is the circumstance  $K$  that *makes*  $S$  fictional, the fiction taken for granted. Letting  $F^K = S$  mean that  $K$  is that circumstance, we want

$$(A') \text{ realcontent}(S) = [\text{the } K \text{ such that } F^K = S].$$

As before, the exponent to which  $F$  needs to be raised to obtain  $S$  can be written as a logarithm:

$$(B') \text{ realcontent}(S) = \log_F(S).$$

This is the defining formula of *object fictionalism*.  $S$  as we use it really means that  $\log_F(S)$ , where  $\log_F(S)$  is not the fact that makes  $S$  true, but the fact that makes it fictional.

Which fact this is depends on the governing fiction, of course. The governing fiction of applied arithmetic says that whenever there are some  $E$ 's, there is an entity *their number* that measures them cardinality-wise; if there are five  $E$ 's, this further entity is 5, if there are a million, it is 1,000,000. The governing fiction of possible worlds theory says that whenever something is possible, there is a world where it happens. The governing fiction of property theory says that whenever there are some  $Q$ 's and nothing else is  $Q$ , there is a property  $Q$ -ness exemplified by all and only those things.<sup>8</sup> Assuming fictions of this general sort, we have

$$\begin{aligned} \log_F(\text{the number of } E\text{'s} = n) &= \textit{there are } n \textit{ } E\text{'s}. \\ \log_F(\text{there is a world such that } H) &= \textit{possibly } H. \\ \log_F(x \text{ has } Q\text{-ness}) &= x \textit{ is } Q. \end{aligned}$$

<sup>8</sup> There will, of course, be more to the fictions than is indicated here.

(B') now tells us that

- realcontent(the number of E's =  $n$ ) = *there are  $n$  E's*.
- realcontent(there is a world such that H) = *possibly H*.
- realcontent( $x$  has Q-ness) =  *$x$  is Q*.

If it is the real content that one really asserts, then

- quasi-asserting "the number of E's =  $n$ " is really asserting *there are  $n$  E's*.
- quasi-asserting "there is an H-world" is really asserting that *possibly H*.
- quasi-asserting " $x$  has Q-ness" is really asserting that  *$x$  is Q*.

In all these cases we rely on the fact that what is true in a story does or can depend on what is true in reality. One gives voice to the real truth by making as if to assert the fictional truth that it enables.

How does this help? Start with the problems of phenomenology and concern. When we say that the number of starving people is large, the real content is that there are many starving people. When we say that the number is rising, the real content is that there were so many starving people yesterday, more today, more tomorrow, and so on. These are facts not about the story but about human beings. And they are facts it makes sense to feel concern about.

Object fictionalism also helps with the modal problem. The real content of " $2 + 3 = 5$ " is the worldly fact that makes it true in the number-story. If the story takes the expected sort of shape, what makes it true in the story that  $2 + 3 = 5$  is that if there are two F's, and three G's, then barring overlap there are five (F-or-G)'s. When this is written as a sentence of first-order logic (numerical quantifiers are defined inductively in the manner of Frege), it is seen to be a logical truth.<sup>9</sup> No wonder " $2 + 3 = 5$ " strikes us as necessary and a priori; at the level of real content, it is.

## 6. THE BOMB

Object fictionalism is on the right track. But, as stated, it is subject to a knock-down objection (the Bomb). It is an objection that arose first in connection with Gideon Rosen's "modal fictionalism" and was generalized to other sorts of fictionalism by Daniel Nolan and John Hawthorne.<sup>10</sup>

S is quasi-assertible iff it is true according to the story; and it is quasi-assertible iff its real content obtains. Consider what this means in the context of applied arithmetic. We have on the one hand that

- (a) " $\#(K's) = n$ " is quasi-assertible iff according to the number-story,  $\#(K's) = n$ .

#(K's)' should be together on the same line; can the whole thing be on one line?

<sup>9</sup> Field has a good discussion in his 1980.

<sup>10</sup> Nolan and O'Leary-Hawthorne 1996.

We have on the other hand that

(b) “ $\#(K's) = n$ ” is quasi-assertible iff there really are  $n$   $K$ 's.

Now, certainly the following is true:

(c) according to the number-story,  $\#(\text{even primes}) = 1$ .

From (a) and (c) it follows that

(d) “ $\#(\text{even primes}) = 1$ ” is quasi-assertible.

From (b) and (d) it follows that

(e) there really is an even prime number.

But of course (e) is not something your typical fictionalist would want to accept; it certainly is not something she wants to be forced into accepting. The whole motivation, after all, was to find a construal of number-talk that did not find you to be actually committed to the things! So it really is a disaster if from the fictionalist's own proposal it follows that numbers do exist—if, as Rosen puts it, the fictionalist winds up a platonist *malgré lui*.

Suppose we turn the argument around, working backwards from the fictionalist's desired result. She wants to maintain (or to preserve the right to maintain) that

(e') there really are no numbers.

From this it follows via (b) that

(d') “ $\#(\text{numbers}) = 0$ ” is quasi-assertible.

This and (a) give us

(c') according to the number-story,  $\#(\text{numbers}) = 0$ .

And that is false. According to the number-story, the number of numbers is very much *larger* than 0. You get the same sort of problem with other applications of the object-fictionalist strategy. Property-fictionalists, for instance, want to be able to say that

(e'') there are no properties.

But then they are committed via (a) and (b) to

(c'') according to the property-story, *being a property* has no instances.

It is not true, though, that according to the property-story, *being a property* has no instances. According to the property-story, *being a property* has lots of instances, namely all the properties.

The results we are getting seem in fact to be worse than false. Take “the number of numbers is 0.” That has no chance of being true, because it is self-refuting. If

the number of numbers is 0, then there is one number at least, namely 0, and so the number of numbers is not 0 after all. Likewise the statement that is said to be true according to the property-story. How could *being a property* have no instances, when it is itself an instance? It seems that either the fictionalist is a platonist, or the story becomes incoherent.

## 7. HOW I LEARNED TO LOVE THE BOMB

“The number of numbers is 0” seems at first obviously false. But there are settings where statements of the same basic form strike us as not false but true. Example: Sometimes we say that a person is “full of it.” (I will understand this as in a familiar way elliptical for something we would rather not dwell on.) Holocaust deniers are full of it; Pat Buchanan is full of it. “Full of it” is, I assume, *never* meant, or taken, literally. But we can imagine a context where this happens. Imagine a speaker Ned so naïve as to think that when Buchanan is described as full of it, this is to be understood as a surprising but well-supported claim about the contents of Buchanan’s body. A joker Jerry has been feeding this information to Ned, and Ned has come to believe that Jerry is right. Jerry has told Ned that when people (including Jerry himself) describe Buchanan as full of it, they are to be taken literally. What can we say to Ned to set him straight?

One thing we might say is that Buchanan is not really or literally full of it. Ned objects that his friend Jerry has told him otherwise. Why, if Buchanan is not really full of it, is this so often said? It is *not* often said, we explain; “full of it” is a figure of speech. Ned replies that this begs the question. It seems a figure of speech to us, because we think a literal interpretation would be uncharitable, because we do not appreciate the true facts about Buchanan as expounded by Jerry. At this point we are likely to just throw up our hands and say

(?#%) anyone who says people are full of it is full of it!

Now, try if you can to forget the preamble and look at the last sentence again. Taken out of context, it *looks* self-defeating in the manner of “the number of numbers is 0.” It looks like you are calling a certain kind of person a liar, a kind that includes you yourself. But, of course, there is another way of hearing (?#%) so that it makes perfect sense. If we take the first occurrence of “full of it” literally, and the second occurrence figuratively, it says that people who say that people are *literally* full of it are liars. And that is true.

Suppose that Ned is naïve not just about “full of it” but also about “the number of E’s is so and so.” He has been told by Jerry that the point and purpose of saying that the number of Martian moons is 2 is to state an identity between one number, the one that numbers the Martian moons, and another, the number 2. If we are nominalistically inclined fictionalists, we will tell Ned that Jerry’s

interpretation is problematic, because there aren't any numbers. He replies, but Jerry says there *are*; Jerry says that the number of numbers is huge (aleph-nought). Our counter-reply is that far from being aleph-nought, the number of numbers is 0.

## 8. REFLEXIVE FICTIONALISM

Now we have seen how to say “the number of numbers is 0” and have it come out not self-defeating but, by the nominalist's lights, true. What we haven't seen is what exactly is going on in these cases. What is wrong with the object fictionalist's idea that we introduce X's (numbers, say) to help us to talk about Y's (concreta)?

There are actually two roles X's can play. Sometimes they function as *representational aids*. This is how butterflies function in “I had butterflies in my stomach,” and numbers function in “the number of Martian moons is 2.” Other times they function as *things-represented*. This is how butterflies function in “the butterflies were splattered all over the windscreen,” and how numbers function in “there are no numbers, that's just a way of talking.”

Object fictionalism as written cannot handle this distinction. Object fictionalism never contemplates for a moment that X's will function as things-represented. No surprise then that mechanically applying its rules in cases where they do so function leads to unwanted results.

Reflexive fictionalism is object-fictionalism modified to take account of all this. X's can be representational aids, or not (two possibilities). X's can be things represented, or not (once again, two possibilities). Multiplying two by two, this gives us four types of statement—three, if we leave aside the case in which X's function in neither way (“the cat is on the mat”). Each of the three blocks the self-defeat argument (a)–(e) in a different way.

- (1) There are sentences in which X's function just as representational aids (“the number of Martian moons is 2”). Call this *applied X-talk*. If we are engaged in applied X-talk, then K is a predicate of “regular” things (concreta), not a predicate of X's (numbers). The inference from (b) to (c) has K a predicate of numbers. So if applied X-talk is our game, the argument does not get beyond (b).
- (2) There are sentences in which X's function just as things represented; for instance, “there are numbers” (spoken by the platonist) and “there are no numbers” (spoken by the nominalist) and “there is an even prime” as it occurs in the self-defeat argument. Call this *explicit X-talk*. If I am speaking explicitly, then I am not using numbers as representational aids. But then (a) and (b) are off limits, for they are principles of quasi-assertion, and there is quasi-assertion only when numbers are playing a representational role.

- (3) There are sentences in which X's function in *both* ways, for instance, "the number of even prime numbers is 0" as spoken by a nominalist. 0 is functioning here as a representational aid, while the primes are things-represented. Call this *self-applied* X-talk. If it is self-applied X-talk we are going in for, then (a) is applicable in principle but will be regarded by the nominalist as false. (a) says that quasi-assertibility is truth according to the number fiction. But although it is true in the fiction that the number of even primes is 1, the nominalist (as just noted) quasi-asserts that the number of even primes is 0.

I said that object fictionalism mishandles the distinction between representational aid and thing-represented. It is not blind to the distinction, for it puts numbers in the first category and concreta in the second. The point it misses is that X's can travel back and forth between the two categories. Not only can they change sides between games; they can do it within a game, indeed within a single sentential move.

Reflexive fictionalism tries to take account of the fact that X-sentences are open to multiple interpretations, corresponding to the various ways of divvying up their X-ish allusions between the sincere and the as-if. The obstacle to multiple contents is our all-purpose governing fiction F. So F will have to go. In its place we put make believe games G, where it is understood that different such games can be played with the same sentences and will be as the occasion demands. Changes in the game we are playing with S make for changes in real content, according to the following rule:  $\text{realcontent}(S) = \log_G(S)$ , where G is the operative game. (Applied X-talk becomes self-applied when acceptability in G falls under the control of facts about X's.) The real content is the condition, whatever it is, to which S owes its acceptability in the game.

## 9. RELATIVE REFLEXIVE FICTIONALISM

Now I want to argue that reflexive fictionalism is not much use and to be treated as a station on the way to something better. Imagine that you are a nominalist taking reflexive fictionalism out for a spin. You are excited by the reports you have heard of using X's to talk about X's and have been looking forward to the opportunity to try it yourself. What sort of description shall you attempt first? What about the numbers lends itself to numerical representation? Immediately your excitement begins to fade. You are not in a mood to attempt *any* description of the numbers, because in your view there are no such things. Or, rather, you are not in a mood to attempt any description beyond "there aren't any." It is true that one *can* say that with numbers (as in the last section). But it is hard to see why anyone would bother. "There aren't any" sums the matter up nicely.

Perhaps then we should have made you a *platonist* taking reflexive fictionalism out for a spin. You do not deny the existence of numbers; a subtle transcendental argument persuades you that they are real. It is just that you don't think ordinary people are *talking* about these transcendently motivated entities when they say in ordinary contexts that  $2 + 3 = 5$ . (They are talking about numbers as much or little as one is talking about Shirley's petard—let's say she has one—when describing her as hoist upon it.) There are *some* people, however, namely platonistic philosophers like yourself, who do talk about numbers, and in them reflexive fictionalism would seem to have found its constituency. Believers in numbers presumably want to say useful and informative things about them. And so they should be interested in technologies that help with this project; and the fiction of numbers is just such a technology. You should be excited, then, about what the fictionalist has to offer.

Or should you? You do indeed want to use numbers for representational purposes. But what possible advantage could as-if numbers have over real ones: the numbers that you as a platonist genuinely believe in? Nominalists (you will say) might benefit from the fiction of numbers, since they have no other access to the referential/quantificational maneuvers that numbers enable. But you as a platonist do not need the fiction to engage in these maneuvers. You have been quantifying over numbers all along, as you are entitled to do given your belief that they are there. It is interesting, perhaps, to realize that the fiction could step in if your genuine numbers proved to be an illusion; the same representational advantages would accrue. But even that should not impress you much. If numbers proved to be an illusion, then the descriptive challenges they seemed to present would have proved illusory, too. "There aren't any" is not all that hard to say.

Who then is reflexive fictionalism really benefiting? No one, it seems. And so we need to make changes. There is one last distinction that needs to be folded into the mix.

Suppose that you as a nominalist say "there are not many even prime numbers." There are two ways you might want to be understood. Perhaps you are trying to portray the numbers as they are. You think there aren't any numbers and conclude from this that there aren't any, or hence many, even prime numbers. In this case, we will say you are speaking in a *disengaged* manner.

The other possibility is that you are trying to portray the numbers as they are supposed to be imagined by players of the relevant game. You say "there are not many even primes" because you know that the numbers are to be thought of as including just one even prime. In this case, we will say you are speaking in an *engaged* manner.

If you are a disengaged nominalist, then you and the platonist do not have much to talk about. She insists there are lots of prime numbers, you insist there are none, and the discussion breaks down. When an engaged nominalist meets a platonist, things go better. The platonist says, "Do you think that the number

of even primes is 1?” And you reply, not, “What do you mean, there are no numbers,” but, “Yes, and here is the proof.”

Both of these conversations have (erstwhile) representational aids taking on the role of things-represented. Both, then, should be grist for the reflexive fictionalist mill. Only one of the conversations, though, is allowed by that doctrine. Provision has been made for talking about the numbers as they are, but not for talking about them as they are to be imagined. The disengaged nominalist gets what he wants, but the engaged one is stiffed.

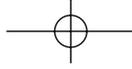
Any real nominalist will want to be both of these characters. He will want to be disengaged when speaking to philosophers (“there being no numbers at all, the number of even primes is 0”), and engaged when doing mathematics (“4, 6, 8, . . . being composite, the number of even primes is 1”). This is a kind of relativism, and so reflexive fictionalism modified to allow for it will be called *relative* reflexive fictionalism. No changes are needed on the disengaged side, but the engaged nominalist (the one who rejects numbers but not number theory) needs our help.

The solution is to allow a new type of game.  $G$  is *basic* if acceptability in  $G$  is a function of how things really are; these are the games we have been talking about so far.  $G^*$  is *parasitic* if acceptability in  $G^*$  depends on how things are imagined to be when playing some other game (as it might be,  $G$ ). The engaged nominalist is speaking parasitically. He is playing a  $G^*$  in which numbers are assigned to the entities imagined to exist when playing  $G$  (= the “applied” game in which numbers are assigned to entities that are really there). The numbers as they are imagined to be in  $G$  include just one that is even and prime;  $G^*$  assigns 1 to the so-and-so’s iff there is only a single so-and-so among the numbers as they are imagined to be in  $G$ ; hence  $G^*$  assigns 1 to the even primes. (The appearance of stratification here is actually somewhat misleading, for parasitic games tend to swallow their hosts; instead of two games, one parasitic on the other, we wind up with a single game parasitic on itself.<sup>11</sup>)

## 10. COMMUNICATION

Nominalists and platonists do not disagree about the number of even primes. Of course, the nominalist is talking about numbers as they are postulated to be in the game. But the platonist has, or should have, no real objection to this. The numbers *she* postulates are supposed to be really there. But apart from that

<sup>11</sup> I am relying here on a perhaps-too-subtle distinction between (i) saying  $S$  meaning: *in the game,  $S$* , and (ii) saying it meaning:  *$S$  (pssst—judge this by its faithfulness to reality as we are supposed to imagine it when doing arithmetic)*. (i) makes life simpler. Our subject matter is always the same: the world as it is. But (ii) better captures how it *feels* to say that there are infinitely many primes. I feel myself to be talking not about the practice, but the objects. (Why otherwise is the infinity of the primes necessary and a priori?) (Thanks to Mark Crimmins for pressing me on this.)



one detail, they are indistinguishable from the numbers as the arithmetic-doing nominalist imagines them.

Saying this goes a *little* way toward addressing an under-discussed problem in the philosophy of mathematics. How is it that mathematicians can happily communicate despite having different views of the nature, and even the existence, of mathematical objects? How can the ontological questions that philosophers sweat over be so irrelevant to actual practice?

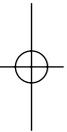
One answer goes back to Carnap. Talk about numbers is internal to the number-framework, and responsible only to that framework's rules of assertion. "True" is a label we apply to the sentences the rules let us utter, and "agreement" is the label we use when the same sentence is uttered by more than one person. Not many have found this answer satisfying. We ask how mathematicians can agree despite their ontological differences, and we are told, in effect, to ignore those differences. To ignore them is to ignore what the discussants take themselves to be saying, and to that extent what they *are* saying. An account of agreement that ignores that is bound to be superficial. It is good that mathematicians do not come to blows over what sentences to accept. But we cannot speak of agreement until we know what the sentences mean in their mouths.

Fictionalism tries to speak to this question. It is agreed how the numbers are to be conceived; they are to be conceived as an omega-sequence generated from 0 by successive application of  $+1$ . And it is agreed that entities answering to that conception would have to have such and such features, such as including infinitely many primes. True, the platonist thinks that there really *are* some things with these features; and perhaps also that we are to conceive the numbers with *these* features because they are the features they really have. True, the nominalist tends to doubt these claims. But that makes little difference in practice. Each hears the other as talking about numbers as they are *taken* to be. They can agree that number-talk is answerable to that, while agreeing to disagree on whether the taking is veridical.

Everyone should now be happy, it seems. The platonist (for whom the taking *is* veridical) gets *real (platonistic) truth*. And the nominalist gets *real agreement* along the lines just sketched. If there is a problem here, it is that the first "real" does not work the same way as the second. Both "real"s look back to real content, but they draw on different aspects of the notion, which we have until recently been running together.

Question: What makes real content real? Answer #1: It concerns *real things*, for instance, moons as opposed to numbers. Call this the *objectual reality* of real content. Answer #2: It is *really asserted*. Call this the *assertional reality* of real content.

One way of putting the moral of the last section is that objectual and assertional reality can come apart. Suppose that I as a nominalist declare the number of primes to be infinite. Assertional reality is not lacking. There is something that I am really saying, as opposed to just pretending to say. But there is no real content





in the objectual sense. I am talking about the numbers as they are supposed to be imagined, not the numbers as they are. This has consequences for the communication problem.

The platonist agrees with the nominalist about the numbers as they are taken to be. But that is not, for her, the content of interest. When she says “there are infinitely many primes,” she means to be claiming that some bona fide numerical entities are infinite in number. The nominalist’s real content is only assertorically real, but hers, she believes, is real in both of our senses. She may then balk at the suggestion that she and the nominalist agree about the number of primes. The nominalist says: you believe what I mean by “there are infinitely many primes,” so we agree. The platonist says: you do not believe what I mean by “there are infinitely many primes,” so we do not agree. (I will not try to address this problem here, except to say that I doubt that the platonist is advancing an objectually real content, as claimed. One does not assert the reality of numbers except when doing philosophy, and we were talking about agreement in mathematics.)

## 11. FIGURALISM

Revolutionary nominalists want us to stop talking about so-and-so’s; hermeneutic nominalists maintain that we never started (Burgess and Rosen 1997). A nominalism based on the methods sketched here would, one imagines, be revolutionary. But then it seems fair to object that the view is extremely complex. A revolution with this many rules is unlikely to generate a whole lot of fervor.

Reply: Actually, the proposal is put forward in more of a hermeneutic spirit. “What, we are all relative reflexive fictionalists without realizing it?” No, but we are something closely related. We are people apt on occasion to speak figuratively. There is nothing in relative reflexive fictionalism not found already in figurative speech.

If, as is sometimes supposed, “linguistic usage [is] literalistic in its main body and metaphorical in its trimming . . . ,” then we are barking up the wrong tree. Applied arithmetic and the like are not special or unusual; they are part of the main body. If, on the other hand, it is non-figurative speech that is special—if, as Quine says, “Cognitive discourse at its most dryly literal is largely a refinement. . . . It is an open space in the tropical jungle, created by clearing tropes away” (1978, 188–9)—then figurative fictionalism might be just what the doctor ordered. Already we have seen how figures can help. It seemed at first quite mysterious how “the number of even primes is zero” could possibly be true—and still more mysterious how it could *also* be correct (as in mathematical contexts it seems to be) to say that the number of even primes is *not* zero. Then a metaphor was found with the same convoluted-looking structure. “Those calling people full of it are full of it” is true when the first “full of it” is taken literally, otherwise false.



“Full of it” is, if I may so put it, the tip of the iceberg. A lot of the phenomena that fictionalists are called on to explain are present in figurative speech and handled effortlessly there. Some examples:

- (A) “7 is less than 11”  
 $\approx$  “the back burner is kept at a lower temperature than the front”;  
 “pinpricks of conscience register less than pangs of conscience”; “a molehill is smaller than a mountain”
- (B) “11 is prime”  
 $\approx$  “the back burner is where things are left to simmer”; “the invisible hand operates all by itself, without encouragement or supervision”;  
 “emotions run high when the green-eyed monster visits”
- (C) “prime numbers are mostly odd”  
 $\approx$  “stomach-butterflies do not sit still but flutter about”; “apron-strings are short”; “molehills are nothing to get excited about”
- (D) “the number of F’s is large iff there are many F’s”  
 $\approx$  “your marital status changes iff you get married or . . .”; “your identity is secret iff no one knows who you are”; “your prospects improve iff it becomes likelier that you will succeed”
- (E) “the F’s outnumber the G’s iff  $\{x|Fx\}$  is bigger than  $\{x|Gx\}$ ”  
 $\approx$  “they are more audacious than you iff they have more gall”; “those are more widely available than these iff their market penetration is greater”; “they are better justified than you iff their reasons are better”
- (F) “the # of F’s = the # of G’s iff there are as many F’s as G’s”  
 $\approx$  “our greatest regret = yours iff we most regret that so-and-so and so do you”; “our level of material well-being = yours iff we are equally well off”; “my bottom line is the same as yours iff both of us are prepared to settle for such-and-such and neither is prepared to settle for anything less”

The similarities here run deep. All of the above statements seem *necessarily true*. It is no accident that if there are as many F’s as G’s, then the F’s and G’s have the same number. It is no accident that if neither of us is prepared to settle for less than the other, then our bottom lines are the same.

Second, all of the statements employ a distinctive vocabulary—“number,” “bottom line”—that can also be used to make *contingent claims about concrete reality* (“the number of sheep exceeds the number of goats,” “negotiations have been difficult, because their bottom line keeps on changing”).

Third, its suitability for making contingent claims about concrete reality is the vocabulary’s *reason for being*. No one cares about stomach-butterflies as such; the question of interest is whether people have butterflies in their stomach. Just so, our interest in 11 has less to do with its relations to 7 than with whether, say, the eggs in a carton have 11 as their number, and what that

means about the carton's relation to other cartons whose eggs have a different number.

Fourth, the vocabulary's utility for this purpose *does not depend* on conceiving of its referential-looking elements as truly standing for things. Those, if any, who take bottom lines and numbers dead seriously derive the exact same expressive benefit from them as those who adamantly deny their existence. And both of these groups derive the exact same expressive benefit as those who never gave the matter the slightest thought.

## 12. RATIONALE

At one time the rationale for fictionalism was obvious. We had, or thought we had, good philosophical arguments to show that X's did not exist, or could not be known about if they did. X's were *obnoxious*, so we had to find an interpretation of our talk that did not leave us committed to them.

That form of argument is dead and gone, it seems to me. It requires very strong premises about the sort of entity that can be known about, or that can plausibly exist; and these premises can always be exposed to ridicule by proposing the numbers themselves as paradigm-case counterexamples.<sup>12</sup>

But there is another possible rationale for fictionalism. Just maybe, it gives the most plausible account of the practice. It is not that X's are intolerable, but that when we examine X-language in a calm and unprejudiced way, it turns out to have a whole lot in common with language that is fictional on its face. If one now asks which elements of everyday speech are fictional on their face, the answer is the figurative elements.

I can illustrate with, you guessed it, the example of numbers. The decision between platonism (including here platonistic semantics) and fictionalism (the figuralist variety) turns on four related questions.

+PLA What does platonism help us to explain? What phenomena are there that make more sense if platonism is true?

–PLA What explanatory puzzles does platonism generate? What becomes hard to make sense of if platonism is true?

+FIG What does figuralism help us to explain? Are there phenomena that make more sense if figuralism is true?

–FIG What explanatory puzzles does figuralism generate? What is there in the figuralist picture that seems puzzling or inexplicable?

Arguments from obnoxiousness ignore +PLA, +FIG, and –FIG to focus on –PLA. One is left to suppose that platonism comes out so far ahead on the other questions that the figuralist needs a big win on –PLA to survive.

<sup>12</sup> Burgess and Rosen 1997.

This is where old-style fictionalism makes its big mistake. It allows and even encourages the notion that the benefits are all on the side of platonism, and that the only way to oppose platonism is to harp on the terrible costs. A better strategy is to say that the “benefits” are largely nonexistent, and the figuralist can explain more than you thought, on a less fanciful basis than you thought. I cannot argue these points in detail here, but some examples will give the flavor.

### What Does Platonism Help Us to Explain?

If there really are numbers, then there is an objective fact of the matter about which arithmetical statements are true. Take the numbers away, and all that is left is the human practice of developing and swapping around proofs, plausibility arguments, suggestive analogies, etc. And that practice, not to say it isn’t highly disciplined, cannot provide as objective a basis for arithmetical truth as a bona fide number series would. The decision problem for arithmetic is of staggering complexity. There is nothing *we* can do to decide matters this complex. That is a task for the numbers themselves.

*Response:* Either our conception of the numbers is determinate or it is not. By “determinate,” I mean that for any arithmetical claim *S*, one of the following is determinately correct: (i) any structure *N* answering to our conception would be such that *S*, or (ii) any structure *N* answering to our conception would be such that not-*S*. Our conception is indeterminate if there are arithmetical claims *S* such that an *N* answering to our conception might or might not be such that *S*. Our conception leaves certain things open which, settled one way, make for an *S*-structure, and settled another way, make for a structure such that not-*S*.

Suppose first that our conception of the numbers is determinate. Then the numbers are not needed for objectivity. Our conception draws a bright line between true and false, whether anything answers to it or not.

If our conception is *not* determinate, there is a question as to how we nevertheless manage to pick out the intended structure. (I assume that we do pick it out, up to isomorphism, since if not, even the platonist has objectivity problems.)

The answer has got to be that the world meets us halfway; of the various technically eligible candidates, only one exists. “The numbers” are whatever out there best corresponds to our not fully determinate intentions.

This, however, makes it a (conceptually) contingent matter which arithmetical claims are correct. There will be arithmetical claims *S* that are true in our mouths, but false in the mouths of our intrinsic duplicates—false in the mouths of (conceptually) possible people just like us internally but who live in a universe with undetectably different numbers.

Arithmetical concepts are not supposed to be externalist in this way. It should not be that although I am right when I say that there are infinitely many primes differing by two, my doppelganger on Twin-Prime Earth is wrong when he says

the same thing. If there are infinitely many twin primes, the reason should not be that such and such are the number-like entities that happen to exist.

So the number-hypothesis, conceived as objectivity-bolstering, is faced with a dilemma. If we are clear enough about what we mean by it, then the hypothesis is not *needed* for objectivity. And if we are not clear what we mean, then it is not going to help. It is not even going to be tolerable, because arithmetical truth is going to blow with the ontological winds in a way that nobody wants.

### What Does Figuralism Help Us to Explain?

*Insubstantiality.* Numbers are thin; they lack (in Mark Johnston’s phrase) a “hidden substantial nature.” There is no more to them than the concept of a number demands. Even if we are not able to work out all that that entails, we do know some of the features that are *not* entailed, and the suggestion that these nevertheless apply strikes us as comical. All of this is what you would expect of something conjured up for representational purposes. Why should we have filled out the story further than needed?

*Indeterminacy.* Numbers’ identity-relations are strikingly less determinate than those of regular objects. There are *lots* of things  $X$  such that there fails to be a fact of the matter as to whether  $7 = X$ . This is only natural if 7 is made up. (There is no fact of the matter either as to whether my keister = my wazoo, or the chip on my shoulder today = the one that was there yesterday.)

*Translucency.* You “see through” my statement that the number of zebra mussels has doubled in a year to the fact I was trying to get across: there are twice as many zebra mussels as a year ago. You do not even register the as-if reference to numbers, and you are surprised when it is pointed out. This makes sense if numbers are representational aids rather (or more) than things-represented. (You “see through” my statement that Gandhi had a lot of guts in the same way.)

*Impatience.* People making statements purporting to be about numbers are strangely indifferent to the question of their existence. Suppose that you as a math teacher tell Fred that what 2 and 3 add up to is 5. And suppose some meddler points out that according to the Oracle (which let us assume we all trust), everything is concrete and so not a number. Instead of calling Fred in to confess your mistake, you tell the meddler to bug off. This makes sense if the meddler’s information is irrelevant to what you were really saying—as indeed it is if your message was that it is *five* things (not six as Fred had supposed) that two things and three other things amount to. (Compare being rebuked for suggesting that Gandhi had a mind of his own on the basis that Gandhi is wholly physical.)

*Representationality.* All abstract objects yet discovered have “turned out” to come in handy as representational aids. How is this interesting coincidence to be explained? Why have numbers, sets, properties, and so on all turned out to be liable to the same sort of use? This should remind us (says the

figuralist) of Wittgenstein's fable in which we first invent clocks, and only later realize that they could be used to tell time. It is no big surprise if things with representing as their reason for "being" show a consistent aptitude for the task.

*Necessity and A Priority.* That a thing should *exist* is the paradigm of a contingent, a posteriori, state of affairs. Yet arithmetic, which is up to its neck in existential commitments, strikes us as a priori and necessary. Why? Suppose as suggested above that the real content of "2 + 3 = 5" is  $(\exists_2x Fx \ \& \ \exists_3y Gy \ \& \ \neg \exists z (Fz \ \& \ Gz)) \rightarrow \exists_5u (Fu \vee Gu)$ . This is a logical truth, and to that extent an a priori necessity. Arithmetic seems necessary and a priori because properly understood, it is.<sup>13</sup>

### What Explanatory Puzzles Does Figuralism Generate?

If we are just pretending to assert, when we say that the number of planets is 9, shouldn't we know it? How does the figuralist propose to explain our obliviousness on this score?

One form of the objection has already been discussed: we are *not* just pretending to assert, when we say the number of planets is 9. We are really asserting that there are nine planets. But the claim will be that we are not pretending *at all*: not even instrumentally as a way of asserting something believed.

If pretending is *making believe*, where "making" signifies an act deliberately undertaken, then the objection seems right. Nothing like that happens when we exchange notes on the number of planets.

But does the figuralist need it to happen? Making believe is an amalgam of (i) being as if you believe, and (ii) being that way through your deliberate efforts. It is only (i) that the figuralist needs. Call it *simulation*.<sup>14</sup> Someone is simulating belief that S if although things are in relevant respects *as if* they believed that S, when they reflect on the matter they find that they do not believe it; or at least are agnostic on the matter; or at least do not feel the propriety of their stance to depend on their belief that S if they have one. They do not believe that S *except possibly per accidens*.

Simulating is being in relevant respects as if one believed, while not believing except possibly per accidens. Copernicus after realizing the astronomical facts still simulates belief in a setting sun. Einstein having developed relativity theory still simulates belief in absolute rest and motion. A movie-goer who realizes full well she is looking at moving images may still simulate the belief that she is being attacked by a giant squid. A dreamer may simulate the belief that she is winning the Nobel Prize.<sup>15</sup>

<sup>13</sup> See "Abstract Objects: A Case Study."

<sup>14</sup> I take the term, or this way of using it, from Walton, "Spelunking, Simulation, and Slime."

<sup>15</sup> The examples are from Walton; he provides a lot of interesting detail.

Making believe is a conscious activity, or one easily brought to consciousness. Simulating is not. It may even come as a great *surprise* that one is simulating. It came as a great surprise to me to realize that although it was as if I believed that an invalid argument was one with countermodels, I did not *really* believe it save per accidens—for I did not believe in models save per accidens.

Someone who utters a sentence committed to X's is to that extent simulating belief that X's exist; for uttering that sentence is a way (not always a deep or thoroughgoing way) of bringing oneself into a relation of resemblance with the (possibly hypothetical) person who believes the sentence's literal content.

### 13. SUMMING UP

The predicament we started with can be stated as follows: what are our options when we discover that we are (only) simulating a belief in X's? As before, one option is to stop simulating by ceasing to be as if a believer in X's. A second option is to stop simulating by becoming a genuine believer in X's. (Or, in light of the "except per accidens" clause, one should come to *express* genuine belief in uttering the sentence.) A third is to keep on simulating, but only when one is in possession of a paraphrase that one can really believe. (Or, in light of the "except per accidens" clause, really believe qua utterer of that sentence.)

What the fictionalist offers is a fourth option. Your simulated beliefs and assertions may be tracking a realm of genuine facts, or a realm of what you take to be facts. If so, then it becomes tempting to construe the simulated assertion that S as a real assertion about the relevant facts—the facts that make a simulation like that appropriate.

There is a danger, though, of that construal seeming contrived or self-serving. Am I engaged in legitimate self-analysis, or am I "tampering with the record" to bring past statements in line with current beliefs? This is where the specifically figuralist version of the doctrine comes in. One answers the charge of self-servingness by pointing out how *similar* our talk of X's is to our (certifiably figurative) talk of Y's.

If Hattie says, "The prof put a lot of hurdles in my path," it is not at all contrived to regard her as simulating to some small extent the belief that her professor (literally) put a lot of hurdles in her path. And it is not at all contrived to regard her as *really* expressing a belief to the effect that her professor made it in thus and such ways difficult for her to accomplish what she had wanted to. The challenge in any particular case is to make out that one's talk of X's *resembles* figurative speech enough to make this sort of construal ring true.

Deciding whether a construal "rings true" is a difficult task, not made easier by our tendency toward wishful thinking and the rewriting of history. It may be that figurism is a tool inherently liable to a certain sort of misuse. One certainly

hopes for more and better controls on the operation than I have been able to provide in this paper.

But a tool liable to misuse is not automatically worthless. It may even be indispensable for some purposes. Compare the notion of conversational implicature. Grice came to regret his invention to some extent; he was not sure he knew how to use it responsibly, that is, non-opportunistically. He never concluded, though, that one should scrap the idea. Implicature happens, so there is no real option but to try to develop a working relationship with it. Figuration happens, too. You learn by trying.

#### REFERENCES

- Alston, W. 1958. "Ontological Commitment." *Philosophical Studies* 9: 8–17.
- Balaguer, M. 1996. "A Fictionalist Account of the Indispensable Applications of Mathematics." *Philosophical Studies* 83: 291–314.
- 1998. *Platonism and Anti-Platonism in Mathematics*. New York: Oxford University Press.
- Bratman, M. 1992. "Practical Reasoning and Acceptance in a Context." *Mind* 101: 1–15.
- Burgess, J. and G. Rosen. 1997. *A Subject with No Object*. Oxford: Clarendon Press.
- Carnap, R. 1956. "Empiricism, Semantics, and Ontology," in his *Meaning and Necessity*, 2nd edition. Chicago: University of Chicago Press.
- Cohen, J. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon.
- Crimmins, M. 1998. "Hesperus and Phosphorus: Sense, Reference, and Pretence." *Philosophical Review* 107: 1–47.
- Davidson, D. 1977. "The Method of Truth in Metaphysics." *Midwest Studies in Philosophy* 2: 244–254.
- 1978. "What Metaphors Mean," in Sacks 1979.
- Davies, M. 1983. "Idiom and Metaphor." *Proceedings of the Aristotelian Society* 83: 67–86.
- Davies, M. and T. Stone. 1995. *Mental Simulation*. Oxford: Blackwell.
- Derrida, J. 1982. "White Mythology: Metaphor in the Text of Philosophy." In *Margins of Philosophy*. Chicago: University of Chicago Press.
- Field, H. 1980. *Science without Numbers*. Princeton: Princeton University Press.
- 1989. *Realism, Mathematics, and Modality*. Oxford: Blackwell.
- Fodor, J. 1964. "On Knowing What We Would Say." *Philosophical Review* 73: 198–212.
- Fodor, J. and J. Katz. 1963. "The Availability of What We Say." *Philosophical Review* 72: 57–71.
- Gibbs, R. W. 1994. *The Poetics of Mind: Figurative Thought, Language, and Understanding*. New York: Cambridge University Press.
- Hills, D. 1998. "Aptness and Truth in Metaphorical Utterance." *Philosophical Topics* 25: 117–153.
- Katz, A., C. Cacciari, R. Gibbs, and M. Turner. 1998. *Figurative Language and Thought*. New York: Oxford University Press.

- Lakoff, G. and R. E. Nuñez. *Where Mathematics Comes From*. New York: Basic Books.
- Lewis, D. K. 1988. "Statements Partly about Observation." *Philosophical Papers* 17:1–31.
- Maddy, P. 1997. *Naturalism in Mathematics*. Oxford: Clarendon.
- Melia, J. 1995. "On What There's Not." *Analysis* 55: 223–229.
- Nolan, D. and J. O'Leary-Hawthorne. 1996. "Reflexive Fictionalisms." *Analysis* 56: 23–32.
- Ortony, A. 1993. *Metaphor and Thought*. 2nd edition. Cambridge: Cambridge University Press.
- Quine, W. V. 1948. "On What There Is." *Review of Metaphysics* 2, reprinted in Quine 1953.
- 1953. *From a Logical Point of View*. Cambridge: Harvard University Press.
- 1960. *Word and Object*. Cambridge: MIT Press.
- 1978. "A Postscript on Metaphor." In Sacks 1978.
- Rosen, G. 1990. "Modal Fictionalism." *Mind* 99: 327–354.
- 1993. "A Problem for Fictionalism about Possible Worlds." *Analysis* 53: 71–81.
- Sacks, S., ed. 1978. *On Metaphor*. Chicago: University of Chicago Press.
- van Fraassen, B. 1980. *The Scientific Image*. Oxford: Oxford University Press.
- Velleman, D. 2000. "On the Aim of Belief." In *The Possibility of Practical Reason*. Oxford: Clarendon.
- Walton, K. 1990. *Mimesis and Make-Believe*. Cambridge: Harvard University Press.
- 1993. "Metaphor and Prop Oriented Make-Believe." *European Journal of Philosophy* 1: 39–57.
- 1997. "Spelunking, Simulation, and Slime." In M. Hjort and S. Laver, eds., *Emotion and the Arts*. Oxford: Oxford University Press.
- 2000. "Existence as Metaphor." In A. Everett and T. Hofweber, eds., *Empty Names, Fiction, and the Puzzles of Existence*. Stanford: CSLI.
- Wright, C. 1983. *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.
- Yablo, S. 1996. "How in the World?" *Philosophical Topics* 24: 255–286.
- 1998. "Does Ontology Rest on a Mistake?" *Proceedings of the Aristotelian Society* supp. vol. 72: 229–262 [Chapter 5 in this volume].
- 2000. "Apriority and Existence." In P. Boghossian and C. Peacocke, eds., *New Essays on the A Priori*. Oxford: Oxford University Press [also in this volume]. **chapter?**
- 2002. "Abstract Objects: A Case Study." *Philosophical Issues* 12:220–240 [Chapter 8 in this volume].
- 2005. "The Myth of the Seven" in M.E. Kalderon, ed., *Fictionalism in Metaphysics* Oxford: Clarendon Press [Chapter 9 in this volume].



## 8

# Abstract Objects: A Case Study

### 1. NECESSITY

Not a whole lot is essential to me: my identity, my kind, my origins, consequences of these, and that is pretty much it. Of my intrinsic properties, it seems arguable that none are essential, or at least none specific enough to distinguish me from others of my kind. And, without getting into the question of whether existence is a property, it is certainly no part of my essence to exist.

I have by contrast *huge* numbers of accidental properties, both intrinsic and extrinsic. Almost any property one would ordinarily think of is a property I could have existed without.



So, if you are looking for an example of a thing whose “essence” (properties had essentially) is dwarfed by its “accense” (properties had accidentally), you couldn’t do much better than me. Of course, you couldn’t easily do much *worse* than me, either. Accense dwarfs essence for just about any old object you care to mention: mountain, donkey, cell phone, or what have you.

Any old *concrete* object, I mean. Abstract objects, especially *pure* abstracta like 11 and the empty set, are a different story. I do not know what the intrinsic properties of the empty set are, but odds are that they are mostly essential. Pure sets are not the kind of thing we expect to go through intrinsic change between one world and another. Likewise integers, reals, functions on these, and so on.<sup>1</sup>

The pattern repeats itself when we turn to relational properties. My relations to other concrete objects are almost all accidental. But the number 11’s relations to other abstract objects (especially other numbers) would seem to be essential.

The most striking differences have to do with existence. Concrete objects (with the possible exception of “the world,” on one construal of that phrase) are one and all contingent. But the null set and the number 11 are thought to exist

I am grateful to a number of people for criticism and advice; thanks above all to Gideon Rosen, Kit Fine, Gilbert Harman, Mario Gomez-Torrente, Marian David, Ted Sider, Paul Horwich, and Stephen Schiffer.

<sup>1</sup> Although on the Frege-Russell definition of number, there is, arguably, intrinsic change. The empty set can change too, if as Lewis suggests it is definable as the sum of all concreta. But I am talking about what we *intuitively* expect, and no one would call these definitions intuitive.

in every possible world. This is *prima facie* surprising, for one normally supposes that existence is inversely related to essence: the bigger  $x$ 's essence, the "harder" it is for  $x$  to exist, and so the fewer worlds it inhabits. And yet here is a class of objects extremely well endowed in the essence department, and missing from not even a single world.

You would have to be in a coma not to wonder what is going on here. Why is it that so much about abstract objects is essential to them? What is it about numbers *et al.* that makes it so hard for them not to exist? And shouldn't objects that turn up under all possible conditions have impoverished essences as a result?

It may be that I have overstated the phenomenon. Not everyone agrees that numbers even exist, so it is certainly not agreed that they exist necessarily. There would be more agreement if we changed the hypothesis to: numbers exist necessarily provided they *can* exist, that is, unless they're impossible.<sup>2</sup> And still more if we made it: numbers exist necessarily provided they *do* exist. But these are nuances and details. I think it is fair to say that *everyone*, even those who opt in the end for a different view, has trouble with the idea that 11 could go missing.

So our questions are in order, construed as questions about how things intuitively seem. Why should a numberless world seem impossible (allowing that the appearance may be only *prima facie*)? Why should it seem impossible for numbers to have had different intrinsic properties, or different relational properties *vis-à-vis* other abstract objects? Why should numbers seem so modally inflexible?

## 2. APRIORITY

A second *prima facie* difference between the concrete and abstract realms is epistemological. Our knowledge of concreta is aposteriori. But our knowledge of numbers, at least, has often been considered apriori. That  $3 + 5 = 8$  is a fact that we *could* know on the basis of experience—of counting, say, or of being told that  $3 + 5 = 8$ . But the same is true of most things we know apriori. It is enough for apriority that experience does not *have* to figure in our justification. And this seems true of many arithmetical claims. One can determine that  $3 + 5 = 8$  just by thinking about the matter.

<sup>2</sup> Wright & Hale suggest in "Nominalism and the Contingency of Abstract Objects" that Field might not accept even that much. Field *does* say that numbers are conceptually contingent. But it would be hard to pin a metaphysical contingency thesis on him, for two reasons. (1) He is on record as having not much use for the notion of metaphysical necessity. (2) To the extent that he tolerates it, he understands it as conceptual entailment by contextually salient metaphysical truths. If salient truths include the fact that *everything is concrete*, then (assuming they are not concrete) numbers will come out metaphysically impossible.

Like the felt necessity of arithmetic, its felt apriority is puzzling and in need of explanation. It is a thesis of arithmetic that there are these things called numbers. And it is hard to see how one could be in a position to know apriori that things like that really existed.

It helps to remember the two main existence-proofs philosophers have attempted. The ontological argument tries to deduce God's existence from God's definition, or the concept of God. The knock against this has been the same ever since Kant; from the conditions a thing would have to satisfy to be X, nothing existential follows, unless you have reason to think that the conditions are in fact satisfied. Then there is Descartes's cogito. This could hardly be expected to give us much guidance about how to argue apriori for numbers. Also, the argument is not obviously apriori. You need to know that you think, and that knowledge seems based on your experience of self.<sup>3</sup>

I said that the ontological argument and the cogito were the two best-known existence-proofs in philosophy. Running close behind is Frege's attempted derivation of numbers themselves. If the Fregean line is right, then numbers are guaranteed by logic together with definitions. Shouldn't that be enough to make their existence apriori? Perhaps, if the logic involved were ontology-free. But Frege's logic affirms the existence of all kinds of higher-type objects.<sup>4</sup> (Frege would not have wanted to *call* them objects because they are not saturated; but there is little comfort in that.) The Fregean argument cannot defeat doubts about apriori existence, because it presupposes they *have* been defeated in presupposing the apriority of Fregean logic.

A different strategy for obtaining apriori knowledge of numbers goes via the "consistency-truth principle": in mathematics, a consistent theory is a true theory. If we can know apriori that theory T is consistent, and that the consistency-truth principle holds, we have apriori warrant for thinking T is true, its existential claims included.

There are a lot of things one could question in this strategy. Where do we get our knowledge of the consistency-truth principle? You may say that it follows from the fact that consistent theories have (intended) models, and that truth is judged relative to those models. But that argument assumes the truth of model theory. And apriori knowledge of model theory does not seem easier to get than apriori knowledge of arithmetic.

Even if we do somehow know the consistency-truth principle apriori, a problem remains. Not all consistent theories are on a par. Peano Arithmetic, one feels, is *true*, and other theories of the numbers (AP) are true only to the extent that they agree with PA. It doesn't help to say that PA is true of its portion of mathematical reality, while AP is true of its. That if anything only reinforces the

<sup>3</sup> Burge (2000) takes a different view (p. 28).

<sup>4</sup> See Rayo and Yablo (2006) for an interpretation that (supposedly) frees the logic of these commitments.

problem, because it makes AP just as true in its own way as PA. It begins to look as though arithmetical truth can be apriori only if we downgrade the kind of truth involved. A statement is not true/false absolutely but only relative to a certain type of theory or model.

### 3. ABSOLUTENESS

I take it as a given that mathematical truth doesn't *feel* relative in this way. It feels as though  $3 + 5$  is just plain 8. It feels as though the power set of a set is just plain bigger than the set itself.

It could be argued that the notion of truth at work here is still at bottom a relativistic one: it is truth according to *standard math*, where a theory is standard if the mathematical community accepts and uses it.<sup>5</sup>

But truth-according-to-accepted-theories is a far cry from what we want, and act like we have. For now the question becomes, why is this theory standard and not that? The answer cannot be that the theory is *true*, in a way that logically coherent alternatives are not true, because there is no truth on this view but truth-according-to-accepted-theories; to explain acceptance in terms of acceptance-relative truth would be to explain it in terms of itself. I assume then that PA's acceptance will have to be traced to its greater utility or naturalness given our projects and cognitive dispositions. But this has problematic results. Why is it that  $3 + 5 = 8$ ? Because we wound up *passing* on the coherent alternative theory according to which  $3 + 5$  is not 8—and for reasons having nothing to do with truth. Neither theory is truer than the other. That, as already stated, is not at all how it feels.

Another problem is sociological. 3 plus 5 was seen to be 8 long before anyone had formulated a theory of arithmetic. How many people even today know that arithmetic is something that mathematicians have a theory of? Saul and Gloria (my non-academic parents) are not thinking that  $3 + 5 = 8$  is true-relative-to-the-standard-theory, because they have no idea that such a theory exists, and if apprised of it would most likely think that the theory was standard because it was true. Are they just confused? If so, then someone should pull the scales from their eyes. Someone should make them realize that the truth about numbers and sets is (like the truth about what's polite or what's stylish) relative to an unacknowledged standard, a standard that is in relevant respects quite arbitrary. I would not want to attempt it, and not only because I don't like my parents angry at me. If they would balk at the notion that there's no more to be said for standard mathematics than for a successful code of etiquette, I suspect they're probably right.

<sup>5</sup> See Balaguer (2001) for discussion.

Admittedly, there are *parts* of mathematics, especially of set theory, where a relative notion of truth seems not out of place. Perhaps the most we can say about the continuum hypothesis is that in some nice-looking models it is true, while in others it is false. I admit, then, that the intuition of absolute truth may not extend to all cases. But even in set theory it extends pretty far. A set theory denying, say, Infinity, or Power Set, strikes us as *wrong*, even if we have yet to put our finger on where the wrongness is coming from.

Could the explanation be as simple as this? If a model doesn't satisfy Power Set, or Infinity, then we don't see it as modeling "the sets." That Infinity holds in all models of "the sets" is a trivial consequence of that linguistic determination. It's not as if there is a shortage of models which include only *finite* set-like objects. It's just that these objects are at best the pseudo-sets, and that makes them irrelevant to the correctness of Infinity taken as a description of the sets. Infinity is "true" because models that threaten to falsify it are shown the door; they are not part of the theory's intended subject matter.

Call this the *debunking* explanation of why it seems wrong to deny the standard axioms. I do not say that the debunking explanation is out of the question; it may be that ZF serves in effect as a reference-fixer for "set." But again, that is not how it feels. If someone wants to argue that Infinity is wrong—that the hereditarily finite sets are the only ones there are—our response isn't "save your breath! deny Infinity and you're changing the subject." Our response is: "that sounds unlikely, but let's hear the argument." No doubt we will end up thinking that the Infinity-denier is wrong. The point is that what he is wrong about is *the sets*. It *has* to be, for if he is not talking about the sets, then we are not really in disagreement.

Suppose, though, the debunkers are right that ZF is true because it sets the standard for what counts as a set. This still doesn't quite explain the feeling that ZF is in some absolute sense correct. Why should we be so obsessed with the *sets* as opposed to the pseudo-sets defined by theory FZ? To the extent that ZF and "sets" are a pair, curiosity about why ZF seems so right is a lot like curiosity about why the sets seem so right. It doesn't matter how the questions individuate, as long as they're both in order. And so far nothing has been said to cast doubt on this. So again, why do ZF and the sets seem so right?

#### 4. ABSTRACTNESS AND NECESSITY

Three puzzles, then: one about necessity, one about apriority, one about absoluteness. It will be easiest to start with necessity; the other two puzzles will be brought in shortly.

The necessity puzzle has to do both with essential properties and necessary existence. About the latter it may be speculated that there is something about *abstractness* that prevents a thing from popping in and out of existence as we

travel from world to world.<sup>6</sup> It is, as Hale and Wright put it,<sup>7</sup> hard to think what conditions favorable for the emergence of numbers would be, and hard to think of conditions unfavorable for their emergence. It is by contrast easy to think of conditions favorable for the emergence of Mt. McKinley. The reason, one imagines, is that numbers are abstract and Mt. McKinley is not.

But, granted that numbers do not wait for conditions to be right, how does that bear on their necessity? Explanations come to an end somewhere, and when they are gone we are left with the brute facts. Why shouldn't the existence/nonexistence of numbers be a brute fact? Traditionally existence has been the paradigm of a phenomenon not always admitting of further explanation. Granted that numbers are not contingent *on* anything, one still wants to know why they should not be contingent full stop.

A second possible explanation is that it is part of the *concept* of an abstract object (a "pure" abstract object, anyway) to exist necessarily if at all. An object that appeared in this world but not others would by that alone not be abstract.

Suppose that is right; an otherwise qualified object that does not persist through all worlds does not make the cut. One might still be curious about these contingent would-be abstracta. What sort of object are we talking about here? The obvious thought is that they are *exactly like real abstracta* except in the matter of necessary existence. But the obvious thought is strange, and so let us ask explicitly: Could there be shabstract objects that are just like their abstract cousins except in failing to persist into every world?

Fiddling with an object's persistence conditions is generally considered harmless. If I want to introduce, or call attention to, a kind of entity that is just like a person except in its transworld career—it is missing (e.g.) from worlds where the corresponding person was born in Latvia—then there would seem to be nothing to stop me. If we can have shmersons alongside persons, why not shnumbers along with numbers?

You may think that there is a principled answer to this: a principled reason why abstracta cannot be "refined" so as to exist in not quite so many worlds. If so, though, then you hold the view that we started with: there is something about *abstractness* that precludes contingency. What is it? Earlier we looked at the idea that where pure abstracta like numbers are concerned, there could be no possible basis for selection of one world over another. But why should that bother us? Why should the choice of worlds not be arbitrary, with a different number-refinement for each arbitrary choice? This is only one suggestion, of course, but as far as I am aware, the route from abstractness to necessity has never been convincingly sketched.

<sup>6</sup> Impure abstracta like singleton-Socrates are not thought to exist in all possible worlds. So really I should be talking about pure-abstractness. I'll stick to "abstract" and leave the qualification to be understood. (Thanks here to Marian David.)

<sup>7</sup> Hale and Wright (1996).



Suppose, then, that abstract objects *can* be refined. There is nothing *wrong* with shabstract objects, on this view, it is just that they should not be confused with *abstract* objects. Another set of questions now comes to the fore. Why do we attach so much importance to a concept—abstractness—that rules out contingent existence, as opposed to another—shabstractness—that differs from the first only in being open to contingent existence? Does the salience of numbers as against shnumbers reflect no more than a random preference for one concept over another? One would like to think that more was involved.

## 5. CONSERVATIVENESS AND NECESSITY

So far we have been looking at “straight” explanations of arithmetical necessity: explanations that accept the phenomenon as genuine and try to say why it arises. Attention now shifts to non-straight or “subversive” explanations. Hartry Field does not think there are any numbers. So he is certainly not going to try to *validate* our intuition of necessary existence. He might however be able to *explain the intuition away*, by reinterpreting it as an intuition not of necessity but something related. He does in fact make a suggestion along these lines. Field calls a theory *conservative* if

it is consistent with every internally consistent theory that is ‘purely about the physical world’ (Field 1989, 240).

Conservative theories are theories compatible with any story that might be told about how things go physically, as long as that story is consistent in itself. (I am going to skate lightly over the controversy over how best to understand “consistent” and “compatible” here. The details are not important for what follows.)

Now, one obvious way for a mathematical theory to be conservative is for it to be *necessary*. A theory that cannot help but be true is *automatically* compatible with every internally consistent physical theory.

But, although necessity guarantees conservativeness, there can be conservativeness without it. A necessary theory demands nothing; every world has what it takes to make the theory true. A conservative theory makes no demands on the *physical* world. If the theory is false, it is false not for physical reasons but because the world fails to comply in some other way. T is conservative iff for each world in which T is false, there’s another, physically just like the first, in which T is true. The theory is false, then, only due to the absence of non-physical objects like numbers.

You might think of the foregoing as a kind of necessity. A conservative theory T is “quasi-necessary” in the sense that *necessarily, T is satisfiable in the obtaining physical circumstances*. Here again is Field:



mathematical realists . . . have held that good mathematical theories are not only true but necessarily true; and a clear part of the content of this (the only clear part, I think) is that mathematics is conservative. . . . Conservativeness might loosely be thought of as ‘necessary truth without the truth.’ . . . I think that the only clear difference between a conservative theory and a necessarily true one is that the conservative theory need not be true. . . . Perhaps many realists would be content to say that all they meant when they called mathematical claims necessarily true was that they were true and that the totality of them constituted a conservative theory (Field 1989, 242).

From this it seems a small step to the suggestion that the only distinctively *modal* intuition we have about mathematical objects is that the theory of those objects is conservative. So construed, the intuition is quite correct. And it is correct in a way that sits well with our feeling that existence is never “automatic”—that nothing has such a strong grip on reality as to be incapable of not showing up.

Is our intuition of the necessity of “ $3 + 5 = 8$ ” just a (confused) intuition of quasi-necessity, that is, conservativeness?

I think it is very unlikely. Yes, every world has a physical duplicate with numbers. But one could equally go in the opposite direction: every world has a physical duplicate without them. If the permanent possibility of adding the numbers in makes for an intuition of necessity, then the permanent possibility of taking them out should make us want to call numbers impossible. And the second intuition is largely lacking. A premise that is symmetrical as regards mathematical existence cannot explain why numbers seem necessary as opposed to impossible.

A second reason why necessity is not well-modeled by conservativeness is this. Arithmetical statements strike us as *individually* necessary. We say, “this has *got* to be true,” not “this considered in the context of such and such a larger theory has got to be true.” But the latter is what we *should* say if our intuition is really of conservativeness. For conservativeness is a property of particular statements only seen as exemplars of a surrounding theory. A statement that is conservative in the context of one theory might change stripes in the context of another. (Imagine, for instance, that it is inconsistent with the other.) Nothing like that happens with necessity.

A third problem grows out of the discussion above of consistency as sufficient for truth. Suppose that two theories contradict each other. Then intuitively, they cannot both be necessary; indeed if one is necessary, then the other is impossible. But theories that contradict each other *can* both be conservative.

Someone might reply that if contradictory means *syntactically* contradictory, then contradictory theories can so be necessary. All we have to do is think of them as describing different domains (different portions of the set-theoretic universe, perhaps).

That is true in a technical sense. But the phenomenon to be explained—our intuition of necessity—occurs in a context where contradictory theories are, the technical point notwithstanding, experienced as incompatible. If I affirm Infinity and you deny it, we take ourselves to be disagreeing. But both of us are saying something conservative over physics.

When two statements contradict each other, they cannot both be necessarily true. Unless, of course, the truth is *relativized*: to the background theory, a certain type of model, a certain portion of mathematical reality. This takes us out of the frying pan and into another frying pan just as hot. Once we relativize, standard mathematics ceases to be *right* (full stop). And as already discussed, a lot of it *feels* right (full stop). Once again, then, our problems about apriority and necessity are pushing us toward a no less problematic relativism.

## 6. FIGURALISM<sup>8</sup>

The conservativeness gambit has many virtues, not least its short way with abstract ontology. At the same time there are grounds for complaint. One would have liked an approach that made arithmetic “necessary” without making it in a correlative sense “impossible.” And one would have liked an approach less friendly to relativism.

The best thing, of course, would be if we could hold onto the advantages of the Field proposal without giving up on “real” necessity, and without giving up on the intuition of absolute truth or correctness. Is this possible? I think it just may be. I can indicate the intended direction by hazarding (what may strike you as) some extremely weird analogies:

- (A) “7 is less than 11”
  - “the frying pan is not as hot as the fire”
  - “a molehill is smaller than a mountain”
  - “pinpricks of conscience register less than pangs of conscience”
- (B) “7 is prime”
  - “the back burner is where things are left to simmer”
  - “the average star has a rational number of planets”
  - “the real estate bug doesn’t sting, it bites”
- (C) “primes over two are not even but odd”
  - “butterflies in the stomach do not sit quietly but flutter about”
  - “pounds of flesh are not given but taken”
  - “the chips on people’s shoulders never migrate to the knee”

<sup>8</sup> This section corresponds to section 11 of “Go Figure”.

- (D) “the number of *F*s is large iff there are many *F*s”  
 “your marital status changes iff you get married or . . .”  
 “your identity is secret iff no one knows who you are”  
 “your prospects improve iff it becomes likelier that you will succeed”
- (E) “the *F*s outnumber the *G*s iff  $\# \{x \mid Fx\} > \# \{x \mid Gx\}$ .”  
 “you are more resolute . . . iff you have greater resolve”  
 “these are more available . . . iff their market penetration is greater”  
 “he is more audacious . . . iff he has more gall”
- (F) “the # of *F*s = the # of *G*s iff there are as many *F*s as *G*s”  
 “your whereabouts = our whereabouts iff you are where we are”  
 “our greatest regret = yours iff we most regret that . . . and so do you”  
 “our level of material well-being = yours iff we are equally well off”

Here are some ways in which these statements appear to be analogous. (I will focus for the time being on necessity.)

All of the statements seem, I hope, true. But their truth does not depend on what may be going on in the realm of concrete objects and their contingent properties and relations. There is no way, we feel, that 7 could fail to be less than 11. Someone who disagrees is not understanding the sentence as we do. There is no way that molehills could fail to be smaller than mountains, even if we discover a race of mutant giant moles. Someone who thinks molehills could be bigger is confused about how these expressions work.

Second, all of the statements employ a *distinctive vocabulary*—“number,” “butterflies,” “ $\{x \mid Fx\}$ ,” “market penetration”—a vocabulary that can also be used to talk about concrete objects and their contingent properties. One says “the number of stars is constantly growing,” “his marital status is constantly changing,” and so on.

Third, its suitability for making contingent claims about concrete reality is the vocabulary’s *reason for being*. Our interest in stomach-butterflies does not stem from curiosity about the aerodynamics of fluttering. All that matters to us is whether people *have* butterflies in the stomach on particular occasions. Our interest in 11 has less to do with its relations to 7 than with whether, say, the eggs in a carton have 11 as their number, and what that means about the carton’s relation to other cartons whose eggs have a different number.

Fourth, the vocabulary’s utility for this purpose *does not depend* on conceiving of its referential-looking elements as genuinely standing for anything. It doesn’t depend on conceiving its referential-looking elements any other way, either. Those if any who take stomach-butterflies, greatest regrets, and numbers dead seriously derive the exact same expressive benefit from them as those who think the first group insane. And both groups derive the exact same expressive benefit as the silent majority who have never given the matter the slightest thought.

## 7. NECESSITY AS BACK-PROPAGATED

I said that all of the statements strike us as necessary, but I did not offer an explanation of why. With regard to the non-mathematical statements, an explanation is quickly forthcoming.

Stomach-butterflies and the rest are *representational aids*. They are “things” that we advert to not (not at first, anyway) out of any interest in what they are like in themselves, but because of the help they give us in describing other things. Their importance lies in the way they boost the language’s expressive power.

By making as if to assert that I have butterflies in my stomach, I really assert something about how I feel—something that it is difficult or inconvenient or perhaps just too *boring* to put literally. The *real content* of my utterance is the real-world condition that makes it sayable that S. The real content of my utterance is that reality has feature BLAH: the feature by which it fulfills its part of the S bargain.

The reason it seems contingent that someone’s marital status has changed is that, at the level of real content, it *is* contingent: they could have called the whole thing off. The reason it seems necessary that our prospects have improved iff we are likelier to succeed is that, at the level of real content, it *is* necessary, as the two sides say the very same thing.

How does the world have to be to hold up its end of the “the number of planets is even” bargain? How does the world have to be to make it sayable that the number of planets is even, supposing for argument’s sake that there are numbers? There have to be evenly many planets. So, the real content of “the number of planets is even” is that there are evenly many planets.

That there are evenly many planets is a hypothesis that need not have been true, and that it takes experience to confirm. At the level of real content, then, “the number of planets is even” is epistemically and metaphysically contingent. But there might be *other* number-involving sentences whose real contents are necessary. To the extent that it is their real contents we hear these sentences as expressing, it will be natural for us to think of the sentences as necessarily true.

This explains how number-involving sentences, e.g., “the number of Fs = the number of Gs iff the Fs and Gs are equinumerous” can feel necessary, at the same time as we have trouble seeing how they *could* be necessary. Our two reactions are to different contents. The sentence feels necessary because at the level of real content it is tautologous: the Fs and Gs are equinumerous iff they are equinumerous. And tautologies really are necessary.

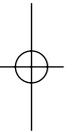
The reason we have trouble crediting our first response is that the sentence’s literal content—that there is this object, a number, that behaves like so—is to the effect that something exists. And it is baffling how anything could cling to existence that tightly.



Why do the two contents get mooshed together in this way? A sentence's *conventional* content—what it is generally understood to say—can be hard to tell apart from its *literal* content. It takes work to remember that the literal meaning of “he’s not the brightest guy in town” leaves it open that he’s the second brightest. It takes work to remember that (literally) pouring your heart out to your beloved would involve considerable mess and a lengthy hospital stay, not to mention the effect on your beloved. Since there is no reason for us to do this work, it is not generally realized what the literal content in fact is.

Consider now “ $7 < 11$ .” To most (!) people, most of the time, it means that seven somethings are fewer than eleven somethings. But the literal content is quite different. The literal content makes play with entities 7 and 11 that measure pluralities size-wise, and encode by their internal relations facts about supernumerosity. Of course, the plurality-measures 7 and 11 are no more on the speaker’s mind than blood is on the mind of someone offering to pour their heart out. “ $7 < 11$ ” is rarely used to describe numbers as such, and so one forgets that the literal content is about nothing else.

The literal contents of pure-mathematical statements are quickly recovered, once we set our minds to it. The real contents remain to be specified. I do not actually think that the real contents are always the same, so there is a considerable amount of exaggeration in what follows. But that having been said, the claim will be that arithmetic is, at the level of real content, a body of logical truths—specifically, logical truths about cardinality—while set theory consists, at the level of real content, of logical truths of a combinatorial nature.



## 8. ARITHMETIC

Numbers enable us to make claims which have as their real contents things we really believe, and would otherwise have trouble putting into words.

One can imagine introducing number-talk for this purpose in various ways, but the simplest is probably this. Imagine that we start out speaking a first-order language with variables ranging over concreta. Numerical quantifiers “ $\exists_n x Fx$ ” are defined in the usual recursive way.<sup>9</sup> Now we adopt the following rule (\*S\* means that it is to be supposed or imagined that S):

(N) if  $\exists_n x Fx$ , then \*there is a thing  $n$  = the number of Fs\*.

Since (N)’s antecedent states the real-world condition under which we’re to make as if the Fs have a number, F should be a predicate of concrete objects. But the reasons for assigning numbers to concrete pluralities apply just as much

<sup>9</sup>  $\exists_0 x Fx =_{\text{df}} \forall x (Fx \rightarrow x \neq x)$ , and  $\exists_{n+1} x Fx =_{\text{df}} \exists y (Fy \ \& \ \exists_n x (Fx \ \& \ x \neq y))$





to pluralities of numbers (and pluralities of both together). So (N) needs to be strengthened to

(N) if  $*\exists_n x Fx*$  then  $*\text{there is a thing } n = \text{the number of } Fs*$ .

This time  $F$  is a predicate of concreta and/or numbers. Because the rule works recursively in the manner of Frege, it gets us “all” the numbers even if there are only finitely many concreta. 0 is the number of non-self-identical things, and  $k+1$  is the number of numbers  $\leq k$ .

Making as if there are numbers is a bit of a chore; why bother? Numbers are there to expedite cardinality-talk. Saying “ $\#Fs = 5$ ” instead of “ $\exists_5 x Fx$ ” puts the numeral in a quantifiable position. And we know the expressive advantages that quantification brings. Suppose you want to get it across to your neighbor that there are more sheep in the field than cows. Pre-(N) this takes (or would take) an infinite disjunction: there are no cows and one sheep or there are no cows and two sheep or there is one cow and there are two sheep, and etc. Post-(N) we can say simply that the number of sheep, whatever it may be, exceeds the number of cows. The real content of “ $\# \text{ sheep} > \# \text{ cows}$ ” is the infinite disjunction, expressed now in finite compass.<sup>10</sup>

This gives a sense of what the real contents of *applied* arithmetical statements are; statements of *pure* arithmetic are another matter.

Take first quantifierless addition statements. What does the concrete world have to be like for it to be the case that, assuming numbers,  $3+5 = 8$ ? Assuming numbers is assuming that there is a number  $k$  numbering the  $Fs$  iff there are  $k$   $Fs$ . But that is not all. One assumes that if no  $Fs$  are  $Gs$ , then the number of  $Fs$  and the number of  $Gs$  have a sum = the number of things that are either  $F$  or  $G$ . Hence, the real-world condition that makes it OK to suppose that  $3+5 = 8$  is that

$$\exists_3 x Fx \ \& \ \exists_5 y Gy \ \& \ \forall x \neg(Fx \ \& \ Gx) \ \rightarrow \ \exists_8 z (Fz \ \vee \ Gz).$$

This is a logical truth. Consider next quantifierless multiplication statements. What does the concrete world have to be like for it to be the case that, assuming numbers,  $3 \times 5 = 15$ ? Well, it is part of the number story that if  $n = \text{the number of } F_1s = \text{the number of } F_2s = \dots = \text{the number of } F_ms$ , and there is no overlap between the  $F_1s$ , then  $m$  and  $n$  have a product  $m \times n = \text{the number of things that are } F_1 \text{ or } F_2 \text{ or } \dots \text{ or } F_m$ . The real-world condition that entitles us to suppose that  $3 \times 5 = 15$  is

$$(\exists_3 x F_1x \ \& \ \dots \ \& \ \exists_3 x F_5x \ \& \ \neg \exists x (F_1x \ \& \ F_2x) \ \& \ \dots) \ \rightarrow \ \exists_{15} x (F_1x \ \vee \ \dots \ \vee \ F_5x).$$

<sup>10</sup> Compare Harry Field’s views on the reason for having a truth-predicate, in the absence of any corresponding property.



Once again, this is a logical truth. Negated addition and multiplication statements are handled similarly; the real content of  $3 + 3 \neq 6$ , for example, is that

$$\exists_3x Fx \ \& \ \exists_5y Gy \ \& \ \forall x \neg(Fx \ \& \ Gx) \ \rightarrow \ \neg \exists_9z (Fz \ \vee \ Gz).$$

CT#

Of course, most arithmetical statements, and all of the “interesting” ones, have quantifiers. Can logically true real contents be found for them?

They can, if we help ourselves to a few assumptions. First, the real content of a universal (existential) generalization over numbers is given by the countable conjunction (disjunction) of the real contents of its instances. Second, conjunctions all of whose conjuncts are logically true are logically true. Third, disjunctions any of whose disjuncts are logically true are logically true. From these it follows that

*The real content of any arithmetical truth is a logical truth.*

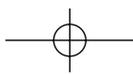
Atomic and negated-atomic truths have already been discussed.<sup>11</sup> These give us all arithmetical truths (up to logical equivalence) when closed under four operations: (1) conjunctions of truths are true; (2) disjunctions with truths are true; (3) universal generalizations with only true instances are true; (4) existential generalizations with any true instances are true. It is not hard to check that each of the four operations preserves the property of being logically true at the level of real content. We can illustrate with case (4). Suppose that  $\exists x \phi(x)$  has a true instance  $\phi(n)$ . By hypothesis of induction,  $\phi(n)$  is logically true at the level of real content. But the real content of  $\exists x \phi(x)$  is a disjunction with the real content of  $\phi(n)$  as a disjunct. So the real content of  $\exists x \phi(x)$  is logically true as well.

## 9. SET THEORY

Sets are nice for the same reason as numbers. They make possible sentences whose real contents we believe, but would otherwise have trouble putting into words. One can imagine introducing set-talk for this purpose in various ways, but the simplest is probably this. “In the beginning” we speak a first-order language with quantifiers ranging over concreta. The quantifiers can be singular or plural; one can say “there is a rock such that it . . .” and also “there are some rocks such that they . . .” Now we adopt the following rule:

- (S) if there are some things  $a, b, c, \dots$ , then \*there is a set  $\{a, b, c, \dots\}$ .\*

<sup>11</sup> Mario Gomez-Torrente pointed out that some atomic truths have not been fitted out with real contents, a fortiori not with logically true real contents. An example is  $(3 + 2) + 1 = 6$ . This had me worried, until he pointed that these overlooked atomic truths were logically equivalent to non-atomic truths that hadn’t been overlooked. For instance,  $(3 + 2) + 1 = 6$  is equivalent to  $\exists y ((3 + 2 = y) \ \& \ (y + 1 = 6))$ . A quick and dirty fix is to think of overlooked sentences as inheriting real content from their not overlooked logical equivalents.



Since the antecedent here states the real-world condition under which we're to make as if  $a, b, c, \dots$  form a set,  $a, b, c, \dots$  are limited to concrete objects. But the reasons for collecting concreta into sets apply just as much to the abstract objects introduced via (S). So (S) is strengthened to

(S) if \*there are some things  $a, b, c \dots$ \*, then \*there is a set  $\{a, b, c, \dots\}$ \*.

This rule, like (N) in the last section, works recursively. On the first go-round we get sets of concreta. On the second go-round we get sets containing concreta and/or sets of concreta. On the third we get sets containing concreta, sets of them, and sets of *them*. And so on through all the finite ranks. Assuming that there are only finitely many concreta, our output so far is the *hereditarily finite* sets: the sets that in addition to being themselves finite have finite sets as their members, and so on until we reach the concrete objects that started us off.

What now? If we think of (S) as being applied at regular intervals, say once a minute, then it will take all of eternity to obtain the sets that are hereditarily finite. No time will be left to obtain anything else, ~~for example,~~ the first infinite number  $\omega$ . ^

The answer to this is that we are not supposed to think of (S) as applied at regular intervals; we are not supposed to think of it as applied at all. (S) does not say that when we *establish* the pretense-worthiness of "there are these things," it *becomes* pretense-worthy that "they form a set." It says that if as a matter of fact (established or not) \*there are these things,\* then \*there is the set of them.\* If \*there are the hereditarily finite sets\*, then certainly \*there are the von Neumann integers ( $0 = \phi, n + 1 = \{0, 1, \dots, n\}$ )\*. And now (S) tells us that \*there is the set  $\{0, 1, 2, 3, \dots\}$ \*, in other words, \*there is  $\omega$ \*. Similar reasoning gets us all sets of rank  $\alpha$  for each ordinal  $\alpha$ . (S) yields in other words the full tower of sets: the full cumulative hierarchy.

Now, to say that (S) yields the full cumulative hierarchy might seem to suggest that (S) yields *a certain fixed bunch* of sets, viz. all of them. That is not the intention. There would be trouble if it were the intention, for (S) leaves no room for a totality of all sets. To see why, suppose for contradiction that \* $a, b, c, \dots$  are all the sets\*<sup>\*</sup>. (S) now tells us that \*all the sets form a set  $V$ \*<sup>\*</sup>. This set  $V$  must for familiar reasons be different from  $a, b, c, \dots$ . So the proposed totality is not all-encompassing. (I will continue to say that (S) yields the full cumulative hierarchy, on the understanding that the hierarchy is not a fixed bunch of sets, since any fixed bunch you might mention leaves something out. This does not prevent a truth-definition, and it does not prevent us from saying that some sentences are true of the hierarchy and the rest false.<sup>12</sup>)

Conjuring up all these sets is a chore; why bother? The reason for bothering with numbers had to do with *cardinality*-type logical truths. Some of these truths

<sup>12</sup> See the last few pages of Putnam (1967) and "Putnam Semantics" in Hellman (1989).



are infinitely complicated, but with numbers you can formulate them in a single finite sentence. Something like that is the rationale for sets as well. The difference is that sets help us to deal with *combinatorial* logical truths—truths about what you get when you combine objects in various ways.

An example will give the flavor. It is a theorem of set theory that if  $x = y$ , then  $\{x, u\} = \{y, v\}$  iff  $u = v$ . What combinatorial fact if any does this theorem encode? Start with “ $\{x, u\} = \{y, v\}$ .” Its real content is that they<sub>xu</sub> are them<sub>yv</sub>—or, to dispense with the plurals, that  $(x = y \text{ or } x = v) \ \& \ (u = y \text{ or } u = v) \ \& \ (y = x \text{ or } y = v) \ \& \ (v = x \text{ or } v = u)$ . Thus what our theorem is really saying is that

If  $x = y$ , then  
 $[(x=y \vee x=v) \wedge (u=y \vee u=v) \wedge (y=x \vee y=v) \wedge (v=x \vee v=u)]$  iff  $u=v$ .

This is pretty simple as logical truths go. Even so it is not really comprehensible; I at least would have trouble explaining what it says. If truths as simple as this induce combinatorial bogglement, it should not be surprising that the set-theoretic formulations are found useful and eventually indispensable.

A second example is Cantor’s Theorem. What is the logical truth here? One can express *parts* of it using the plural quantifier  $\exists X$  (“There are some things such that . . .”). Numerical *plural* quantifiers are defined using the standard recursive trick:

$\exists_0 X \phi(X)$  iff  $\forall X \neg \phi(X)$   
 $\exists_{n+1} X \phi(X)$  iff  $\exists Y (\phi(Y) \ \& \ \exists_n X (\phi(X) \ \& \ X \neq Y))$

Consider now  $\exists_4 X \forall y (Xy \rightarrow Fy)$ . I can’t give this a *very* natural paraphrase, because English does not quantify over pluralities of pluralities. But roughly the claim is that there are four ways of making a selection from the Fs.<sup>13</sup> This lets us express part of what Cantor’s Theorem is “really saying”, viz. that if there are  $n$  Fs, then there are  $2^n$  ways of selecting just some of the Fs, as follows:

$\exists_{n^x} Fx \rightarrow \exists_{2^n} X \forall y (Xy \rightarrow Fy)$ . V

This is a second-order logical truth, albeit a different such truth for each value of  $n$ . But we are still a long way from capturing the Theorem’s real content, because it applies to infinite pluralities as well. There is (as far as I know) no way with the given resources to handle the infinite case.<sup>14</sup> It all becomes rather easy, though, if we are allowed to encode the content with sets. All we need say is that

<sup>13</sup> Alternatively, there are some things all of which are Fs, and some things not the same as the first things all of which are Fs, and etc. (Say there are two Fs. You can pick both of them, either taken alone, or neither of them. Note that “all the Fs” and “none of them” are treated here as limiting cases of “some of the Fs.”)

<sup>14</sup> You could do it with plural quantification over ordered pairs.

the 'n' should be higher than you've got it (it should be above the 2), but not as high as last time; it's super to 2, not to  $\exists$

just like this  $\exists_{2^n} X$

every set, finite or infinite, has more *subsets* than it has *members*. ( $|P(X)| = 2^{|X|} > |X|$ .)

Now let me try to give a general recipe for finding real contents. It will be simplest if we limit ourselves to talk of hereditarily finite sets. Take first atomic sentences, that is, sentences of the form  $x = y$  and  $x \in z$ . A reduction function  $r$  is defined:

- (A<sub>1</sub>)  $r(x \in z)$  is
- |   |                                    |   |
|---|------------------------------------|---|
| 1. $\exists y ((\bigcap_{u \in z} y = u) \ \& \ x = y)$ | if $z$ has members                 | ) |
| 2. $\exists y (y \neq y \ \& \ x = y)$                  | if $z$ is the empty set            | ) |
| 3. $x \in z$  | if $z$ is not a set. <sup>15</sup> |   |

Note that the first line simplifies to  $\bigvee_{y \in z} x = y$ ; that is in practice what I will take the translation to be. (The reason for the quantified version is that it extends better to the case where  $z$  is the empty set.) The third line marks the one place where  $\in$  is not eliminated. If  $z$  is not a set, then it is (literally) false to say that  $x$  belongs to it, which is the result we want. The rule for identity-statements is

- (A<sub>2</sub>)  $r(x = y)$  is
1.  $\forall u (u \in x \leftrightarrow u \in y)$  if  $x$  and  $y$  are sets
  2.  $x = y$  if either is not a set.

In the “usual” case,  $x$  and  $y$  have members, and  $\forall u (u \in x \leftrightarrow u \in y)$  reduces to  $(\bigwedge_{u \in x} \bigvee_{v \in y} u = v) \ \& \ (\bigwedge_{v \in y} \bigvee_{u \in x} v = u)$ . If  $x$  has members and  $y$  is the null set, it reduces to  $\forall u (\bigvee_{z \in x} u = z \leftrightarrow u \neq u)$ . If both  $x$  and  $y$  are the null set, we get  $\forall u (u \neq u \leftrightarrow u \neq u)$ . Otherwise  $r$  leaves  $x = y$  untouched. Non-atomic statements reduce to truth-functional combinations of atomic ones by the following rules:

- (R<sub>1</sub>)  $r(\neg \phi)$  is  $\neg r(\phi)$   
 (R<sub>2</sub>)  $r(\wedge_i \phi_i)$  is  $\wedge_i r(\phi_i)$   
 (R<sub>3</sub>)  $r(\vee_i \phi_i)$  is  $\vee_i r(\phi_i)$   
 (R<sub>4</sub>)  $r(\forall x \phi(x))$  is  $\bigwedge_{z=z} r(\phi(z))$ .  
 (R<sub>5</sub>)  $r(\exists x \phi(x))$  is  $\bigvee_{z=z} r(\phi(z))$ .

The real content of  $\phi$  is found by repeatedly applying  $r$  until you reach a fixed point, that is, a statement  $\phi^*$  such that  $r(\phi^*) = \phi^*$ . This fixed point is a truth-functional combination of “ordinary” statements true or false for concrete (non-mathematical) reasons. These ordinary statements are to the effect that

<sup>15</sup> The idea is that ‘ $x \in z$ ’ describes  $x$  as one of things satisfying the condition for membership in  $z$ . The condition for membership in  $\{a, b, c, \dots\}$  is  $x = a \vee x = b \vee x = c \dots$ . The condition for membership in the null set is  $x \neq x$ .

the spacing  
around '=' is  
looser in A1 than  
A2; A2 is better

⌋  
⌋  
⌋

$x = y$ , where  $x$  and  $y$  are concrete, or  $x = y$ , where one is concrete and the other is not, or  $x \in z$ , where  $z$  is concrete.<sup>16</sup>

How do we know that a fixed point will be reached? If  $\phi$  is a generalization, the  $(R_i)$ s turn it into a truth-functional combination of atoms  $\psi$ . If  $\psi$  is an atom talking about sets, the  $(A_i)$ s turn it into a generalization about sets of a lower rank, and/or non-sets. Now we apply the  $(R_i)$ s again. Given that  $\phi$  contains only finitely many quantifiers, and all the sets are of finite rank, the process must eventually bottom out.<sup>17</sup> The question is how it bottoms out, that is, the character of the sentence  $\phi^*$  that gives  $\phi$ 's real content.

I claim that if  $\phi$  is a set-theoretic truth, then  $\phi^*$  is, not quite a logical truth, but a logical consequence of basic facts about concreta: identity- and distinctness-facts, and facts to the effect that concreta have no members. To have a word for these logical consequences, let's call them *logically true over concrete combinatorics*, or for short *logically true<sub>cc</sub>*. Three assumptions will be needed, analogous to the ones made above for arithmetic. First, the real content of a universal (existential) generalization is given by the countable conjunction (disjunction) of its instances. Second, conjunctions all of whose conjuncts are logically true<sub>cc</sub> are themselves logically true<sub>cc</sub>. Third, disjunctions any of whose disjuncts are logically true<sub>cc</sub> are logically true<sub>cc</sub>.

*Every set-theoretic truth has a logically true<sub>cc</sub> real content.*

The set-theoretic truths (recall that we are limiting ourselves to hereditarily finite sets) are the closure of the atomic and negated-atomic truths under four rules: (1) conjunctions of truths are true; (2) disjunctions with truths are true; (3) universal generalizations with only true instances are true; (4) existential generalizations with any true instances are true. The hard part is to show that atomic and negated-atomic truths are logically true<sub>cc</sub> at the level of real content. The proof is by induction on the ranks of  $x$  and  $y$ .

*Basis Step*

tighten up please

⌋  
⌋

- (a) If  $x$  and  $y$  are concrete, then the real content of  $x = y$  is that  $x = y$ . This is logically true<sub>cc</sub> if true, because it's a consequence of itself. Its negation is logically true<sub>cc</sub> if true for the same reason.
- (b) If  $x$  is concrete and  $y$  is a set, then  $x \neq y$  is true. Its real content  $x \neq y$  is logically true<sub>cc</sub>, because a consequence of the fact that  $x \neq y$ .
- (c) If  $x$  and  $y$  are the null set, then  $x = y$  is true. Its real content  $\forall u (u \neq u \leftrightarrow u \neq u)$  is a logical truth, hence logically true<sub>cc</sub>.

<sup>16</sup> Statements of the first type are necessarily true or necessarily false, depending on whether  $x$  is indeed identical to  $y$ . Statements of the second and third types are necessarily false, since concreta cannot be sets or have members.

<sup>17</sup> The same argument would seem to work with sets of infinite rank; there are no infinite descending chains starting from infinite ordinals either.

- (d) If  $x$  is a non-empty set and  $y$  is the null set, then  $x \neq y$  is true. Its real content  $\neg \forall u (\bigvee_{z \in x} u = z \leftrightarrow u \neq u)$  is logically true, hence logically  $\text{true}_{cc}$ .
- (e) If  $y$  is a non-set then  $x \notin y$  is true. Its real content  $x \notin y$  is logically  $\text{true}_{cc}$  because a consequence of itself.
- (f) If  $y$  is the null set then  $x \notin y$  is true. Its real content  $\neg \exists z (z \neq z \ \& \ x = z)$  is logically  $\text{true}_{cc}$  because logically true.

*Recursion Step*

- (a) If  $x$  and  $y$  are nonempty sets, then  $r(x=y)$  is  $(\bigwedge_{u \in x} \bigvee_{v \in y} u = v) \wedge (\bigwedge_{v \in y} \bigvee_{u \in x} v = u)$ . (a1) If it is true that  $x=y$ , then  $r(x=y)$  is a conjunction of disjunctions, each of which has a true disjunct  $u=v$ . By hypothesis of induction, these true disjuncts have logically  $\text{true}_{cc}$  real contents. So  $r(x=y)$  has a logically  $\text{true}_{cc}$  real content. And the real content of  $r(x=y)$  is also that of  $x=y$ . (a2) If it is true that  $x \neq y$ , then  $r(x \neq y)$  is a disjunction of conjunctions, each of which is built out of true conjuncts. By hypothesis of induction, these true conjuncts are logically  $\text{true}_{cc}$  at the level of real content. So  $r(x \neq y)$  has a logically  $\text{true}_{cc}$  real content. And the real content of  $r(x \neq y)$  is also that of  $x \neq y$ .
- (b) If  $z$  is a nonempty set, then  $r(x \in z)$  is  $\bigvee_{y \in z} x = y$ . (b1) If it is true that  $x \in z$ , this has a true disjunct  $x = y$ . By hypothesis of induction,  $x = y$  has a logically  $\text{true}_{cc}$  real content. But then  $r(x \in z)$  is logically  $\text{true}_{cc}$  at the level of real content, whence so is  $x \in z$ . (b2) If the truth is rather that  $x \notin z$ , then  $r(x \notin z)$  is a conjunction of true conjuncts. By hypothesis of induction, these conjuncts are logically  $\text{true}_{cc}$  at the level of real content. So  $r(x \notin z)$  has a logically  $\text{true}_{cc}$  real content, whence so also does  $x \notin z$ .

10. SUMMING UP

The view that is emerging takes something from Frege and something from Kant; one might call it “Kantian logicism.” The view is Kantian because it sees mathematics as arising out of our representations. Numbers and sets are “there” because they are inscribed on the spectacles through which we see other things. It is logicist because the facts that we see through our numerical spectacles are facts of first-order logic.

And yet the view is in another way the opposite of Kantian. For Kant thinks necessity is imposed *by* our representations, and I am saying that necessity is imposed *on* our representations by the logical truths they encode. Another possible name then is “anti-Kantian logicism.” I will stick with the original name, comforting myself with the notion that the “anti” in “Kantian” can be thought of as springing into semantic action when the occasion demands.

Back now to our three questions. Why does mathematics seem (metaphysically) necessary, and apriori, and absolute? The first and second of these we have answered, at least for the case of arithmetic and set theory. It seems necessary because the real contents of mathematical statements are logical truths. And logical truths really are necessary. It seems apriori because the real contents of mathematical statements are logical truths. And logical truths really are apriori.

That leaves absoluteness. It might seem enough to cite the absoluteness of logical truth; real contents are not logically true relative to this system or that, they are logically true period.

But there is an aspect of the absoluteness question that this fails to address. The absoluteness of logic does perhaps explain why individual arithmetical statements seem in a non-relative sense correct. It does not explain why Peano Arithmetic strikes us as superior to arithmetical theories that contradict it. It does not tell us why the Zermelo-Fraenkel theory of sets strikes us as superior to set theories that contradict it. For it could be that PA is not the only arithmetical theory—ZF is not the only set theory—with the property that its real content is logically true. AP and FZ could be (at the level of real content) just as logically true as PA and ZF. Let me say something about the ZF side of this problem.

If FZ has a logically true real content, it is *not* the content induced by the game sketched above: the game based on principle (S). (Remember, FZ proves some A such that ZF proves  $\neg A$ . Unless something has gone very wrong, A and  $\neg A$  will not come out assertible in the same game.) FZ can be “correct” only if real contents are judged relative to a *different* principle than

if it is to be imagined that there are some things  $x, y, z, \dots$ , then it is to be imagined that there is a set of those things.

This gives us a way out of our difficulties. I said early on that you cannot accuse someone of changing the subject just because they deny some principle of ZF. *But principle (S) is a great deal more basic than anything found in ZF.* If someone has trouble with the idea behind (S)—the idea that when you have got a determinate bunch of things, you are entitled to the *set* of those things—then that person arguably *doesn't* mean the same thing by “set” as we do.<sup>18</sup>

Suppose we call a theory “ZF-like” if it represents the sets as forming a cumulative hierarchy. Then here is an argument that only ZF-like theories get

<sup>18</sup> This might sound funny, given the widespread view that there are *some* things (the sets) that are too many to form a set. This widespread view is at odds with (S) only if it is supposed that there is some definite bunch of things including all and only the sets. If the sets are a definite bunch of things, it is very hard to understand what could be wrong with gathering them together into a further set. I agree with Putnam when he says that “no concrete model [of Zermelo set theory] could be maximal—nor any *nonconcrete* model either, as far as that goes. Even God could not make a model for Zermelo set theory that it would be *mathematically* impossible to extend, and no matter what ‘stuff’ He might use. . . . it is not necessary to think of sets as one system of objects . . . in order to follow assertions about all sets” (1967, 21).

the sets right. If FZ is not ZF-like, then by definition it does not represent sets as forming a cumulative hierarchy. But the cumulative hierarchy comes straight out of (S), the rule that says that if you've got the objects, you've got the set of them as well. So, whatever it is that FZ describes, it is not a system of entities emerging (S)-style out of their members. Emerging (S)-style out of your members is definitive, though, of the sets as we understand them. FZ may well get something right, but that something is not the sets.

## REFERENCES

- Balaguer, Mark. (1996) "A Fictionalist Account of the Indispensable Applications of Mathematics," *Philosophical Studies* 83: 291–314.
- (2001) "A Theory of Mathematical Correctness and Mathematical Truth," *Pacific Philosophical Quarterly* 82: 87–114.
- Burge, Tyler. (2000) "Frege on Apriority," in Paul Boghossian & Christopher Peacocke (eds.), *New Essays on the Apriori* (Oxford: Oxford University Press).
- Burgess, John & Gideon Rosen. (1997) *A Subject with No Object* (Oxford: Clarendon).
- Field, Hartry. (1980) *Science without Numbers* (Princeton: Princeton University Press).
- (1989) *Realism, Mathematics, & Modality* (Oxford: Basil Blackwell).
- Hale, Bob & Crispin Wright. (1996) "Nominalism and the Contingency of Abstract Objects," in M. Schirn (ed.), *Frege: Importance and Legacy* (de-Gruyter: Hawthorne).
- Hellman, Geoffrey. (1989) *Mathematics without Numbers* (Oxford: Clarendon).
- Putnam, Hilary. (1967) "Mathematics without Foundations," *The Journal of Philosophy* 64: 5–22.
- Rayo, Agustin and Stephen Yablo. (2001) "Nominalism Through De-Nominalization," *Noûs* 35: 74–92.
- Walton, Ken. (1993) "Metaphor and Prop Oriented Make-Believe," *European Journal of Philosophy* 1: 39–57.
- Yablo, Stephen. (1996) "How in the World?" *Philosophical Topics* 24: 255–286.
- (1998) "Does Ontology Rest on a Mistake?" *Proceedings of the Aristotelian Society*, supp. vol. 72: 229–262 [Chapter 5 in this volume].
- (2000) "Apriority and Existence," in Paul Boghossian & Christopher Peacocke (eds.), *New Essays on the Apriori* (Oxford: Oxford University Press).
- (2001) "Go Figure: A Path Through Fictionalism," *Midwest Studies in Philosophy* vol. 25 [Chapter 7 in this volume].

in this volume



## 9

### The Myth of the Seven

Mathematics has been called the one area of inquiry that would retain its point even were the physical world to entirely disappear. This might be heard as an argument for platonism: the view that mathematics describes a special abstract department of reality lying far above the physical fray. The necessary truth of mathematics would be due to the fact that the mathematical department of reality had its properties unchangingly and essentially.

I said that it *might* be heard as an argument for platonism, that mathematics stays on point even if the physical objects disappear. However mathematics does not lose its point either if the *mathematical* realm disappears—or, indeed, if it turns out that that realm was empty all along. Consider a fable from John Burgess and Gideon Rosen’s book *A Subject with No Object*:



Finally, after years of waiting, it is your turn to put a question to the Oracle of Philosophy . . . you humbly approach and ask the question that has been consuming you for as long as you can remember: ‘Tell me, O Oracle, what there is. What sorts of things exist?’ To this the Oracle responds: ‘What? You want the whole list? . . . I will tell you this: everything there is is concrete; nothing there is is abstract. . . .’ (Burgess and Rosen, 1997: 3)



Trembling at the implications, you return to civilization to spread the concrete gospel. Your first stop is [your university here], where researchers are confidently reckoning validity in terms of models and insisting on 1-1 functions as a condition of equinumerosity. Flipping over some worktables to get their attention, you demand that these practices be stopped at once. The entities do not exist, hence

I am grateful to Jamie Tappenden, Thomas Hofweber, Carolina Sartorio, Harry Field, Sandy Berkovski, Gideon Rosen, and Paolo Leonardi for comments and criticism. Most of this chapter was written in 1997 and there are places it shows. For one thing, a lot of relevant literature is simply ignored. Also various remarks about the state of the field were truer then than they are now (which is not to say they were particularly true then). My own views have changed too. Where the chapter speaks of ‘*making* as if you believe that *S*’, I would now say ‘*being* as if you believe that *S*, but not really believing it except possibly per accidens’ (see Yablo, 2002a). Related to this, mathematical objects may exist for all I know. I do not rule it out that ‘ $2 + 3 = 5$ ’ is literally true in addition to being metaphorically true, making it a twice-true metaphor along the lines of ‘no man is an island’. I also do not rule it out that ‘ $2 + 3 = 5$ ’ is a maybe-metaphor, to be interpreted literally if so interpreted it is true, otherwise metaphorically. (Compare ‘Nixon had a stunted superego’, to use Jamie Tappenden’s nice example.)



all theoretical reliance on them should cease. They, of course, tell you to bug off and am-scray. Which, come to think of it, is exactly what you yourself would do, if the situation were reversed.

### FREGE'S QUESTION

Frege in *Notes for L. Darmstaedter* asks, 'is arithmetic a game or a science?'<sup>1</sup> He himself thinks that it is a science, albeit one dealing with a special sort of logical object.<sup>2</sup> Arithmetic considered all by itself, just as a formal system, gives, in his view, little evidence of this: 'If we stay within [the] boundaries [of formal arithmetic], its rules appear as arbitrary as those of chess' (*Grundgesetze* II, section 89).<sup>3</sup> The falsity of this initial appearance is revealed only when we widen our gaze and consider the role arithmetic plays in our dealings with the natural world. According to Frege, 'it is applicability alone which elevates arithmetic from a game to the rank of a science' (*Grundgesetze* II, section 91).

One can see why applicability might be thought to have this result. What are the chances of an arbitrary, off the shelf, system of rules performing so brilliantly in so many theoretical contexts? Virtually nil, it seems; 'applicability cannot be an accident' (*Grundgesetze* II, section 89). What else could it be, though, if the rules did not track some sort of reality? Tracking reality is the business of science, so arithmetic is a science.<sup>4</sup>

The surprising thing is that the same phenomenon of applicability that Frege cites in *support* of a scientific interpretation has also been seen as the primary *obstacle* to such an interpretation. Arithmetic qua science is a deductively organized description of *sui generis* objects with no connection to the natural world. Why should objects like that be so useful in natural science = the theory of the natural world? This is an instance of what Eugene Wigner called 'the unreasonable effectiveness of mathematics'.<sup>5</sup>

Applicability thus plays a curious double role in debates about the status of arithmetic, and indeed mathematics more generally. Sometimes it appears as a *datum*, and then the question is, what *lessons* are to be drawn from it? Other times it appears as a *puzzle*, and the question is, what *explains* it, how does it work?

Hearing just that applicability plays these two roles, one might expect the puzzle role to be given priority. That is, we draw such and such lessons because they are the ones that emerge from our story about how applications in fact work.

<sup>1</sup> Beaney (1997: 366).

<sup>2</sup> I am pretending for rhetorical purposes that Frege is still a logicist in 1919.

<sup>3</sup> Geach and Black (1960: 184–7).

<sup>4</sup> He speaks in *Notes for L. Darmstaedter* of 'The miracle of arithmetic'.

<sup>5</sup> See Wigner (1967).

But the pattern has generally been the reverse.<sup>6</sup> The first point people make is that since applicability would be a miracle if the mathematics involved were not true, it's evidence that the mathematics *is* true. The second thing that gets said (what on some theories of evidence is a corollary of the first) is that applicability is *explained* in part by truth. It is admitted, of course, that truth is not the full explanation.<sup>7</sup> But the assumption appears to be that any further considerations will be specific to the mathematics involved and the application.<sup>8</sup> The most that can be said in *general* about why mathematics applies is that it is true.

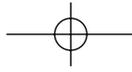
One result of this ordering of the issues is that attention now naturally turns away from applied mathematics to pure. Why should we worry about the bearing of mathematical theories on physical reality, when we have yet to work out their relation to mathematical reality? And so the literature comes to be dominated by a problem I will call *purity*: given that such and such a mathematical theory is true, what makes it true? is arithmetic, for instance, true in virtue of (a) the behavior of particular objects (the numbers), or (b) the behavior of  $\omega$ -sequences in general, or (c) the fact that it follows from Peano's axioms? If (a), are the numbers sets, and if so which ones? If (b), are we talking about actual or possible  $\omega$ -sequences? If (c), are we talking about first-order axioms or second?

Some feel edified by the years of wrangling over these issues, others do not. Either way it seems that something is getting lost in the shuffle, viz., applications. Having served their purpose as a dialectical bludgeon, they are left to take care of themselves. One takes the occasional sidelong glance, to be sure. But this is mainly to reassure ourselves that as long as a mathematical theory is true, there is no reason why empirical scientists should not take advantage of it. That certainly speaks to one possible worry about the use of mathematics in science, namely, is it defensible or something to feel guilty about? But our worry was different:

<sup>6</sup> I am ignoring the Quine/Putnam approach here, first because Quine and Putnam do not purport to draw lessons from *applicability* (but rather indispensability), second because they do not purport to draw *lessons* from applicability. They do not say that we *should* accept mathematics given its applications; they think that we already *do* accept it by virtue of using it, and (this is where the indispensability comes in) we are not in a position to stop.

<sup>7</sup> Asked what had possessed him to drip butter into the Mad Hatter's watch, the Dormouse replied, 'but it was the B E S T butter'. To suppose that truth alone should make for applicability would be like expecting randomly chosen high quality products to improve the operation of randomly chosen machines.

<sup>8</sup> Thus Mark Steiner (1998): 'Arithmetic is useful because bodies belong to reasonably stable families, such as are important in science and everyday life' (25–6). 'Addition is useful because of a *physical* regularity: gathering preserves the existence, the identity, and (what we call) the major properties, of assembled bodies' (27). 'That we can arrange a set [e.g., into rows] without losing members is an empirical precondition of the effectiveness of multiplication . . .' (29). 'Consider now *linearity*: why does it pervade physical laws? Because the sum of two solutions of a (homogeneous) linear equation is again a solution' (30). 'The explanatory challenge . . . is to explain, not the law of gravity by itself, but the prevalence of the inverse square . . . What Pierce is looking for is some general physical property which lies behind the inverse square law, just as the principle of superposition and the principle of smoothness lie behind linearity' (35–6).



Why should scientists *want* to take advantage of mathematics? What good does it do them? What sort of advantage is there to be taken? The reason this matters is that, depending on how we answer, the pure problem is greatly transformed. It could be, after all, that the kind of help mathematics gives is a kind it could give *even if it were false*. If that were so, then the pure problem—which in its usual form presupposes that mathematics is true—will need a different sort of treatment than it is usually given.

### RETOOLING

Here are the main claims so far. Philosophers have tended to emphasize *purity* over *applicability*. The standard line on applicability has been that (i) it is evidence of truth, (ii) truth plays some small role in explaining it, and (iii) beyond that, there is not a whole lot to be said.<sup>9</sup>

A notable exception to all these generalizations is the work of Hartry Field. Not only does Field see applicability as centrally important, he dissents from both aspects of the ‘standard line’ on it. Where the standard line links the utility of mathematics to its truth, Field thinks that mathematics (although certainly useful) is very likely *false*. Where the standard line offers little *other* than truth to explain usefulness, Field lays great stress on the notion that mathematical theories are *conservative* over nominalistic ones, i.e., any nominalistic conclusions that can be proved with mathematics can also be proven (albeit often much less easily) without it.<sup>10</sup> The utility of mathematics lies in the *no-risk deductive assistance that it provides to the beleaguered theorist*.

This is on the right track, I think. But there is something strangely half-way about it. I do not doubt that Field has shown us a way in which mathematics *can* be useful without being true. It can be used to facilitate deduction in nominalistically reformulated theories of his own device: theories that are ‘qualitative’ in nature rather than quantitative. This leaves more or less untouched, however, the problem of how mathematics *does* manage to be useful without being true. It is not as though it benefits only practitioners of Field’s qualitative science (it does not benefit Field-style scientists at all; there aren’t any). The people whose activities we are trying to understand are practicing regular old platonic science.

How without being true does mathematics manage to be of so much help to *them*? Field never quite says.<sup>11</sup> He is quite explicit, in fact, that the relevance of his argument to *actual* applications of mathematics is limited and indirect:

<sup>9</sup> At least, not at this level of generality.

<sup>10</sup> Some have questioned this claim, alleging a confusion of semantic conservativeness with deductive conservativeness. I propose to sidestep that issue entirely.

<sup>11</sup> Field has pointed out to me that there are the materials for an explanation in the representation theorem he proves en route to nominalizing a theory. This is an excellent point and I do not have a worked out answer to it. Let me just make three brief remarks. First, we want an explanation that works even when the theory cannot be nominalized. Second, and more tendentiously, we want





[What I have said] is not of course intended to license the use of mathematical existence assertions in axiom systems for the particular sciences: *Such* a use of mathematics remains, for the nominalist, quite illegitimate. (Or, more accurately, a nominalist should treat such a use of mathematics as a temporary expedient that we indulge in when we don't know how to axiomatize the science properly.) (1980:14)

But then how exactly does he take himself to be addressing our actual situation? I see two main options.

Field might think that the role of mathematics in the *non*-nominalistic theories that scientists really use is *analogous* to its role in connection with his custom-built nominalistic theories—enough so that by explaining and justifying the one, he has explained and justified the other. If that were Field's view, then one suspects he would have done more to develop the analogy.

Is the view, then, that he has *not* explained (or justified) actual applications of mathematics—but that is OK because, come the revolution, these actual applications will be supplanted by the new-style applications of which he *has* treated? This stands our usual approach to recalcitrant phenomena on its head. Usually we try to theorize the phenomena that we find, not popularize the phenomena we have a theory of.

#### INDISPENSABILITY AND APPLICABILITY

As you may have been beginning to suspect, these complaints have been based on a deliberate misunderstanding of Field's project.<sup>12</sup> It is true that he asks:

- (d) What sort of account is possible of how mathematics is applied to the physical world? (Field, 1980, vii)

But this can mean either of two things, depending on whether one is motivated by an interest in *applicability*, or an interest in *indispensability*.

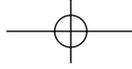
Applicability is, in the first instance, a *problem*: the problem of explaining the effectiveness of mathematics. It is also, potentially, an *argument* for mathematical objects. For the best explanation may require that mathematics is true.

Indispensability is, in the first instance, an *argument* for the existence of mathematical objects. The argument is normally credited to Quine and Putnam. They say that since numbers are indispensable to science, and we are committed to science, we are committed to numbers. But, just as applicability was first a problem, second an argument, indispensability is first an argument, second a

an explanation that doesn't trade on the potential for nominalization even when that potential is there. Third, the explanation that runs through a representation theorem is less a 'deductive utility' explanation than a 'representational aid' explanation of the type advocated later in this paper.

<sup>12</sup> Deliberate now, anyway; it started out as an innocent misunderstanding. Thanks to Ana Carolina Sartorio for straightening me out on these matters.





problem. The problem is: How do nominalists propose to deal with the fact that numbers have a *permanent* position in the range of our quantifiers?

Once this distinction is drawn, it seems clear that Field's concern is more with indispensability than applicability. His question is:

(d-ind) How can applications be conceived so that mathematical objects come out dispensable?

To *this*, Field's two-part package of (i) nominalistically reformulated scientific theories, and (ii) conservation claims, seems a perfectly appropriate answer. But we are still entitled to wonder what Field would say about:

(d-app) How are actual applications to be understood, be the objects indispensable or not?

If there is a complaint to be made, it is not that Field has given a bad answer to (d-app), but that he doesn't address (d-app) at all, and the resources he provides do not appear to ~~be of much use with it.~~ resolve it.

Now, Field *might* reply that the indispensability argument is the important one. But that will be hard to argue. One reason, already mentioned, is that a serious mystery remains even if in-principle dispensability is established. How is the Fieldian nominalist to explain the usefulness-without-truth of mathematics in *ordinary*, quantitative, science? More important, though, suppose that an explanation can be given. Then *indispensability becomes a red herring*. Why should we be asked to *demathematize* science, if ordinary science's mathematical aspects can be understood on some other basis than that they are true? Putting both of these pieces together: The point of nominalizing a theory is not achieved unless a further condition is met, given which condition there is no longer any need to nominalize the theory.

#### NON-DEDUCTIVE USEFULNESS

That is my first reservation about Field's approach. The second is related. Consider the kind of usefulness-without-truth that Field lays so much weight on; mathematics thanks to its conservativeness gives no-risk deductive assistance. It is far from clear why *this particular form* of usefulness-without-truth deserves its special status. It might be thought that there is no other help objects can give without going to the trouble of existing. Field says the following:

if our interest is only with inferences among claims that don't say anything about numbers (but which may employ, say, numerical quantifiers), then we can employ numerical theory without harm, for we will get no conclusions with numerical theory that wouldn't be valid without it . . . There are other purposes for which this justification for feigning acceptance of numerical theory does not apply, and we must decide whether or not to genuinely accept the theory. For instance, there may be observations that we want to



formulate that we don't see how to formulate without reference to numbers, or there may be explanations that we want to state that we can't see how to state without reference to numbers . . . *if such circumstances do arise, then we will have to genuinely accept numerical theory if we are not to reduce our ability to formulate our observations or our explanations* (Field, 1989: 161–2, italics added).

But, *why* will we have to accept numerical theory in these circumstances? Having just maintained that the *deductive* usefulness of *Xs* is not a reason to accept that *Xs* exist, he seems now to be saying that *representational* usefulness is another matter. One might wonder whether there is much of a difference here. I am not denying that deductive usefulness is an important non-evidential reason for making as if to believe in numbers. But it is hard to see why representational usefulness isn't similarly situated.<sup>13</sup>

### NUMBERS AS REPRESENTATIONAL AIDS

What is it that allows us to take our uses of numbers for deductive purposes so lightly? The deductive advantages that 'real' *Xs* do, or would, confer are

<sup>13</sup> Representational usefulness will be the focus in what follows. But I don't want to give the impression that the possibilities end there. Another way that numbers appear to 'help' is by redistributing theoretical content in a way that streamlines theory revision. Suppose that I am working in a first-order language speaking of material objects only. And suppose that my theory says that there are between two and three quarks in each *Z*-particle:

$$(a) (\forall z) [(\exists q_1) (\exists q_2) (q_1 \neq q_2 \ \& \ q_i \in z \ \& \ (\forall r_1) (\forall r_2) ((r_1 \neq r_2 \ \& \ r_j \in z) \rightarrow (r_1 = q_1 \text{ etc.}))].$$

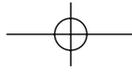
Then I discover that my theory is wrong: The number of quarks in a *Z*-particle is between two and *four*. Substantial revisions are now required in sentence (a). I will need to write in a new quantifier ' $\forall r_3$ '; two new non-identities ' $r_1 \neq r_3$ ' and ' $r_2 \neq r_3$ '; and two new identities ' $r_3 = q_1$ ' and ' $r_3 = q_2$ '. Compare this with the revisions that would have been required had quantification over numbers been allowed—had my initial statement been

$$(b) (\forall z) (\forall n) (n = \#q (q \in z) \rightarrow 2 \leq n \leq 3).$$

Starting from (b), it would have been enough just to strike out the '3' and write in a '4'. So the numerical way of talking seems better able than the non-numerical way to efficiently absorb new information. Someone might say that the revisions would have been just as easy had we helped ourselves to numerical quantifiers ( $\exists_{\geq n} x$ ) defined in the usual recursive way. The original theory numbering the quarks at two or three could have been formulated as

$$(c) (\forall z) [(\exists_{\geq 2} q) q \in z \ \& \ \neg (\exists_{\geq 4} q) q \in z].$$

To obtain the new theory from (c), one need only change the second subscript. But this approach only postpones the inevitable. For our theory might be mistaken in another way: rather than the number of quarks in a *Z*-particle being two or three, it turns out that the number is two, three, five, seven, eleven, or . . . or ninety-seven—that is, the number is a *prime* less than one hundred. If we want to write this in the style of (c), our best option is a disjunction about thirty times longer than the original. Starting from (b), however, it is enough to replace ' $2 \leq n \leq 3$ ' with ' $n$  is prime &  $2 \leq n \leq 100$ .' True, we could do better if we had a primitive 'there exist primely many . . .' quantifier. But, as is familiar, the strategy of introducing a new primitive for each new expressive need outlives its usefulness fairly quickly. The only really progressive strategy in this area is to embrace quantification over numbers.



(Field tells us) equally conferred by  $X$ s that are just 'supposed' to exist. But the same would appear to apply to the representational advantages conferred by  $X$ s; these advantages don't appear to depend on the  $X$ s really existing either. The economist need not believe in the average family to derive representational advantage from it ('the average family has 2.7 bank accounts'). The psychiatrist need not believe in libido or ego strength to derive representational advantage from them. Why should the physicist have to believe in numbers to access new contents by couching her theory in numerical terms?

Suppose that our physicist is studying escape velocity. She discovers the factors that determine escape velocity and wants to record her results. She knows a great many facts of the following form:

- (A) A projectile fired at so many meters per second from the surface of a planetary sphere so many kilograms in mass and so many meters in diameter will (will not) escape its gravitational field.

There are problems if she tries to record these facts without quantifying over mathematical objects, that is, using just numerical adjectives. One is that, since velocities range along a continuum, she will have to write uncountably many sentences, employing an uncountable number of distinct adjectives. Second, almost all reals are 'random' in the sense of encoding an irreducibly infinite amount of information.<sup>14</sup> So, unless we think there is room in English for uncountably many semantic primitives, almost all of the uncountably many sentences will have to be infinite in length. At this point someone is likely to ask why we don't drop the numerical-adjective idea and say simply that:

- (B) For all positive real numbers  $M$  and  $R$ , the escape velocity from a sphere of mass  $M$  and diameter  $2R$  is the square root of  $2GM/R$ , where  $G$  is the gravitational constant.

Why not, indeed? To express the infinitely many facts in finite compass, we bring in numbers as representational aids. We do this despite the fact that what we are trying to get across has nothing to do with numbers, and could be expressed without them were it not for the requirements of a finitely based notation.

The question is whether functioning in this way as a representational aid is a privilege reserved to existing things. The answer appears to be that it isn't. That (B) succeeds in gathering together into a single content infinitely many facts of form (A) owes nothing whatever to the real existence of numbers. It is enough that *we understand what (B) asks of the non-numerical world*, the numerical world

<sup>14</sup> It is not just that for every recursive notation, there are reals that it does not reach; most reals are such that no recursive notation can reach them.



taken momentarily for granted.<sup>15</sup> How the real existence of numbers could help or hinder that understanding is difficult to imagine.

An oddity of the situation is that Field makes the same sort of point himself in his writings on truth. He thinks that ‘true’ is a device that exists ‘to serve a certain logical need’—a need that would also be served by infinite conjunction and disjunction if we had them, but (given that we don’t) would go unmet were it not for ‘true’. No need then to take the truth-predicate ontologically seriously; its place in the language is secured by a role it can fill quite regardless of whether it picks out a property. It would seem natural for Field to consider whether the same applies to mathematical objects. Just as truth is an essential aid in the expression of facts not about truth (there is no such property), perhaps numbers are an essential aid in the expression of facts not about numbers (there are no such things).<sup>16</sup>

#### OUR OPPOSITE FIX

To say it one more time, the standard procedure in philosophy of mathematics is to start with the pure problem and leave applicability for later. It comes as no surprise, then, that most philosophical theories of mathematics have more to say about what makes mathematics true than about what makes it so useful in empirical science.

The approach suggested here looks to be in an opposite fix. Our theory of applications is rough but not non-existent. What are we going to say, though,

<sup>15</sup> This point is also stressed by Balaguer. I first heard it from Gideon Rosen in 1990. He suggested defining the nominalistic content of a math-infused statement  $S$  as the set of worlds  $w$  such that  $w$  is indiscernible in concrete respects from some  $w^*$  where  $S$  is true.

<sup>16</sup> Field does remark in various places that there may be no easy way of detaching the ‘material content’ of a statement partly about abstracta:

the task of splitting up mixed statements into purely mathematical and purely non-mathematical components is a highly non-trivial one: it is done easily in [some] cases [e.g., ‘2 = the number of planets closer than the Earth to the Sun’], but it isn’t at all clear how to do it in [other] cases [e.g., ‘for some natural number  $n$  there is a function that maps the natural numbers less than  $n$  onto the set of all particles of matter,’ ‘surrounding each point of physical space-time there is an open region for which there is a 1–1 differentiable mapping of that region onto an open subset of  $R^4$ .’] (Field, 1989: 235)

He goes on to say that:

the task of splitting up all such assertions into two components is precisely the same as the task of showing that mathematics is dispensable in the physical sciences. (Field, 1989: 235)

This may be true if by ‘mathematics is dispensable’ one means (and Field does mean this) ‘in any application of a mixed assertion. . . a purely non-mathematical assertion could take its place’ (235). But in *that* sense of dispensable—ideological dispensability, we might call it—truth is not dispensable either; there is no truth-less way of saying lots of the things we want to say. It appears that ideological indispensability has *in the case of truth* no immediate ontological consequences. Why then is it considered to argue for the existence of numbers?

about pure mathematics? If the line on applications is right, then one suspects that arithmetic, set theory, and so on are largely untrue. At the very least, then, the problem of purity is going to have to be reconceived. It cannot be: In virtue of what is arithmetic true? It will have to be: How is the line drawn between 'acceptable' arithmetical claims and 'unacceptable' ones? And it is very unclear what acceptability could amount to if it floats completely free of truth.

Just maybe there is a clue in the line on applications. Suppose that mathematical objects 'start life' as representational aids. Some systems of mathematical objects will work better in this capacity than others, e.g., standard arithmetic will work better than a modular arithmetic in which all operations are 'mod  $k$ ', that is, when the result threatens to exceed  $k$  we cycle back down to 0. As wisdom accumulates about the kind(s) of mathematical system needed, theorists develop an intuitive sense of what is the right way to go and what is the wrong way. Norms are developed that take on a life of their own, guiding the development of mathematical theories past the point where natural science greatly cares. The process then begins to feed on itself, as descriptive needs arise with respect to, not the natural world, but *our system of representational aids as so far developed*. (After a certain point, the motivation for introducing larger numbers is the help they give us with the mathematical objects already on board.) These needs encourage the construction of still further theory, with further ontology, and so it goes.

You can see where this is headed. If the pressures our descriptive task exerts on us are sufficiently coherent and sharply enough felt, we begin to feel under the same sort of external constraint that is encountered in science itself. Our theory is certainly answerable to *something*, and what more natural candidate than the *objects* of which it purports to give a literally true account? Thus arises the feeling of the objectivity of mathematics qua description of mathematical objects.

#### SOME WAYS OF MAKING AS IF<sup>17</sup>

I can make the above a bit more precise by bringing in some ideas of Kendall Walton's about 'making as if'. The thread that links as-if games together is that they call upon their participants to pretend or imagine that certain things are the case. These to-be-imagined items make up the game's *content*, and to elaborate and adapt oneself to this content is typically the game's very point.<sup>18</sup> At least one of the things we are about in a game of mud pies, for instance, is to work out who has what sorts of pies, how much longer they need to be baked, etc. At least

<sup>17</sup> This section repeats some of Yablo (1998).

<sup>18</sup> Better, such and such is part of the game's content if 'it is to be imagined. . . *should the question arise*, it being understood that often the question *shouldn't arise*' (Walton, 1990: 40). Subject to the usual qualifications, the ideas about make-believe and metaphor in the next few paragraphs are all due to Walton (1990, 1993).

one of the things we're about in a discussion of Sherlock Holmes is to work out, say, how exactly Holmes picked up Moriarty's trail near Reichenbach Falls, how we are to think of Watson as having acquired his war wound, and so on.

As I say, to elaborate and adapt oneself to the game's content is typically the game's very point. An alternative point suggests itself, though, when we reflect that all but the most boring games are played with *props*, whose game-independent properties help to determine what it is that players are supposed to imagine. That Sam's pie is too big for the oven does not follow from the rules of mud pies alone; you have to throw in the fact that Sam's clump of mud fails to fit into the hollow stump. If readers of 'The Final Problem' are to think of Holmes as living nearer to Windsor Castle than Edinburgh Castle, the facts of nineteenth-century geography deserve a large part of the credit.

A game whose content reflects the game-independent properties of worldly props can be seen in two different lights. What ordinarily happens is that we take an interest in the props because and to the extent that they influence the content; one tramps around London in search of 221B Baker Street for the light it may shed on what is true according to the Holmes stories.

*But in principle it could be the other way around:* we could be interested in a game's content because and to the extent that it yielded information about the props. This would not stop us from playing the game, necessarily, but it would tend to confer a different significance on our moves. Pretending within the game to assert that B L A H would be a way of giving voice to a fact holding *outside* the game: the fact that the props are in such and such a condition, viz., the condition that makes B L A H a proper thing to pretend to assert. If we were playing the game in this alternative spirit, then we'd be engaged not in *content-oriented* but *prop-oriented* make-believe. Or, since the prop might as well be the entire world, *world-oriented* make-believe.

It makes a certain in principle sense, then, to use make-believe games for serious descriptive purposes. But is such a thing ever actually done? A case can be made that it is done all the time—not perhaps with explicit self-identified games like 'mud pies' but impromptu everyday games hardly rising to the level of consciousness. Some examples of Walton's suggest how this could be so:

Where in Italy is the town of Crotona? I ask. You explain that it is on the arch of the Italian boot. 'See that thundercloud over there—the big, angry face near the horizon', you say; 'it is headed this way'. . . . We speak of the saddle of a mountain and the shoulder of a highway. . . . All of these cases are linked to make-believe. We think of Italy and the thundercloud as something like pictures. Italy (or a map of Italy) depicts a boot. The cloud is a prop which makes it fictional that there is an angry face. . . . The saddle of a mountain is, fictionally, a horse's saddle. But our interest, in these instances, is not in the make-believe itself, and it is not for the sake of games of make-believe that we regard these things as props. . . . [The make-believe] is useful for articulating, remembering, and communicating facts about the props—about the geography of Italy, or the identity of the storm cloud. . . . or mountain topography. It is by thinking of Italy



or the thundercloud . . . as potential if not actual props that I understand where Crotona is, which cloud is the one being talked about.<sup>19</sup>

A certain kind of make-believe game, Walton says, can be ‘useful for articulating, remembering, and communicating facts’ about aspects of the game-independent world. He might have added that make-believe games can make it easier to reason about such facts, to systematize them, to visualize them, to spot connections with other facts, and to evaluate potential lines of research. That similar virtues have been claimed for metaphors is no accident, if metaphors are themselves moves in world-oriented pretend games. And this is what Walton maintains. A metaphor on his view is an utterance that represents its objects as being *like so*: the way that they *need* to be to make the utterance ‘correct’ in a game that it itself suggests. The game is played not for its own sake but to make clear which game-independent properties are being attributed. They are the ones that do or would confer legitimacy upon the utterance construed as a move in the game.

#### THE KINDS OF MAKING AS IF AND THE KINDS OF MATHEMATICS

Seen in the light of Walton’s theory, our suggestion above can be put like this: numbers as they figure in applied mathematics are *creatures of existential metaphor*. They are part of a realm that we play along with because the pretense affords a desirable—sometimes irreplaceable—mode of access to certain real-world conditions, viz. the conditions that make a pretense like that appropriate in the relevant game. Much as we make as if, e.g., people have associated with them stores of something called ‘luck’, so as to be able to describe some of them metaphorically as individuals whose luck is ‘running out’, we make as if pluralities have associated with them things called ‘numbers’, so as to be able to express an (otherwise hard to express because) infinitely disjunctive fact about relative cardinalities like so: The number of *F*s is divisible by the number of *G*s.

Now, if applied mathematics is to be seen as world-oriented make-believe, then one attractive idea about *pure* mathematical statements is that:

- (C) They are to be understood as *content-oriented* make-believe.

Why not? It seems a truism that pure mathematicians spend most of their time trying to work out what is true according to this or that mathematical theory.<sup>20</sup> All that needs to be added to the truism, to arrive at the conception of pure mathematics as content-oriented make-believe, is this: the mathematician’s

<sup>19</sup> Walton (1993: 40–1).

<sup>20</sup> The theory might be a collection of axioms; it might be that plus some informal depiction of the kind of object the axioms attempt to characterize; or it might be an informal depiction pure and simple.

interest in working out what is true-according-to-the-theory is by and large independent of whether the theory is thought to be *really true*—true in the sense of correctly describing a realm of independently constituted mathematical objects.<sup>21</sup>

That having been said, the statements of at least *some* parts of pure mathematics, like simple arithmetic, are legitimated (made pretense-worthy) by very general facts about the non-numerical world. So, on a natural understanding of the arithmetic game, it is pretendable that  $3 + 5 = 8$  because if there are three  $\mathbb{F}$ s, and five  $\mathbb{G}$ s distinct from the  $\mathbb{F}$ s, then there are eight  $(\mathbb{F} \vee \mathbb{G})$ s—whence construed as a piece of world-oriented make-believe, the statement that  $3 + 5 = 8$  ‘says’ that if there are three  $\mathbb{F}$ s and five  $\mathbb{G}$ s, etc. For at least some pure mathematical statements, then, it is plausible to hold that:

(W) They are to be understood as *world-oriented* make-believe.

Construed as world-oriented make-believe, every statement of ‘true arithmetic’ expresses a first-order logical truth; that is, it has a logical truth for its metaphorical content.<sup>22</sup> (The picture that results might be called ‘Kantian logicism’. It is *Kantian* because it grounds the necessity of arithmetic in the representational character of numbers. Numbers are always ‘there’ because they are written into the spectacles through which we see things. The picture is *logicist* because the facts represented—the facts we see through our numerical spectacles—are facts of first-order logic.)

There is a third interpretation possible for pure-mathematical statements. Arithmeticians imagine that there are numbers. But this a complicated thing to imagine. It would be natural for them to want a codification of what it is that they are taking on board. And it would be natural for them to want this codification in the form of an *autonomous* description of the pretended objects, one that doesn’t look backward to applications. As in any descriptive project, a need may arise for representational aids. *Sometimes* these aids will be the very objects being described: ‘For all  $n$ , the number of prime numbers is larger than  $n$ .’ Sometimes though they will be *additional* objects dreamed up to help us get a handle on the original ones: ‘The number of prime numbers is  $\aleph_0$ .’

What sort of information are these statements giving us? Not information about the concrete world (as on interpretation (w)); the prime numbers form no part of that world. And not, at least not on the face of it, information about the game (as on interpretation (c)); the number of primes would have been aleph-nought even if there had been no game. ‘The number of primes is  $\aleph_0$ ’ gives information about the prime numbers as they are supposed to be conceived by players of the game.

<sup>21</sup> The intended contrast is with true-according-to-some-other-theory.

<sup>22</sup> See Yablo (2002b).



Numbers start life as representational aids. But then, on a second go-round, they come to be treated as a subject-matter in their own right (like Italy or the thundercloud). Just as representational aids are brought in to help us describe other subject-matters, they are brought in to help us describe the numbers. Numbers thus come to play a double role, functioning both as representational aids and things-represented. This gives us a third way of interpreting pure-mathematical statements:

- (M) They are to be understood as prop-oriented make-believe, with numbers etc. serving *both* as props and as representational aids.

One can see in particular cases how they switch from one role to the other. If I say that ‘the number of primes is  $\aleph_0$ ,’ the primes are my subject-matter and  $\aleph_0$  is the representational aid. (This is clear from the fact that I would accept the paraphrase ‘there are denumerably many primes’.) If, as a friend of the continuum hypothesis, I say that ‘the number of alephs no bigger than the continuum is prime’, it is the other way around. The primes are now representational aids and  $\aleph_0$  has become a prop. (I would accept the paraphrase ‘there are primely many alephs no bigger than the continuum’.)

The bulk of pure mathematics is probably best served by interpretation (M). This is the interpretation that applies when we are trying to come up with autonomous descriptions of this or that imagined domain. Our *ultimate* interest may still be in describing the natural world; our *secondary* interest may still be in describing and consolidating the games we use for that purpose. But in most of pure mathematics, world and game have been left far behind, and we confront the numbers, sets, and so on, in full solitary glory.

## TWO TYPES OF METAPHORICAL CORRECTNESS

So much for ‘normal’ pure mathematics, where we work within some existing theory. If the metaphoricalist has a problem about correctness, it does not arise there; for any piece of mathematics amenable to interpretations (C), (W), or (M) is going to have objective correctness conditions. Where a problem *does* seem to arise is in the context of *theory-development*. Why do some ways of constructing mathematical theories, and extending existing ones, strike us as better than others?

I have no really good answer to this, but let me indicate where an answer might be sought. A distinction is often drawn between *true* metaphors and metaphors that are *apt*. That these are two independent species of metaphorical goodness can be seen by looking at cases where they come apart.

An excellent source for the first quality (truth) without the second (aptness) is back issues of *Reader’s Digest* magazine. There one finds jarring, if not necessarily



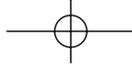
inaccurate, titles along the lines of ‘Tooth Decay: America’s Silent Dental Killer’, ‘The Sino-Soviet Conflict: A Fight in the Family’, and, my personal favorite, ‘South America: Sleeping Giant on Our Doorstep’. Another good source is political metaphor. When Calvin Coolidge said that ‘The future lies ahead’, the problem was not that he was *wrong*—where else would it lie?—but that he didn’t seem to be mobilizing the available metaphorical resources to maximal advantage. (Likewise when George H. Bush told us before the 1992 elections that ‘It’s no exaggeration to say that the undecideds could go one way or another.’)

Of course, a likelier problem with political metaphor is the reverse, that is, aptness without truth. The following are either patently (metaphorically) untrue or can be imagined untrue at no cost to their aptness. Stalin: ‘One death is a tragedy. A million deaths is a statistic.’ Churchill: ‘Man will occasionally stumble over truth, but most times he will pick himself up and carry on.’ Will Rogers: ‘Diplomacy is the art of saying “Nice doggie” until you can find a rock.’ Richard Nixon: ‘America is a pitiful helpless giant.’

Not the best examples, I fear. But let’s move on to the question they were meant to raise. How does metaphorical *aptness* differ from metaphorical *truth*? David Hills (1997: 119–120) observes that where truth is a semantic feature, aptness can often be an aesthetic one: ‘When I call Romeo’s utterance apt, I mean that it possesses some degree of poetic power . . . Aptness is a specialized kind of beauty attaching to interpreted forms of words . . . For a form of words to be apt is for it . . . to be the proper object of a certain kind of felt satisfaction on the part of the audience to which it is addressed.’

That can’t be all there is to it, though; for ‘apt’ is used in connection not just with *particular* metaphorical claims but entire metaphorical frameworks. One says, for instance, that rising pressure is a good metaphor for intense emotion; that possible worlds provide a good metaphor for modality; or that war makes a good (or bad) metaphor for argument. What is meant by this sort of claim? Not that pressure (worlds, war) are metaphorically *true* of emotion (modality, argument). There is no question of truth because no metaphorical claims have been made. But it would be equally silly to speak here of poetic power or beauty. The suggestion seems rather to be that *an as-if game built around pressure (worlds, war) lends itself to the metaphorical expression of truths about emotion (possibility, argument)*. The game ‘lends itself’ in the sense of affording access to lots of those truths, or to particularly important ones, and/or in the sense of presenting those truths in a cognitively or motivationally advantageous light.

Aptness is *at least* a feature of prop-oriented make-believe games; a game is apt relative to such and such a subject-matter to the extent that it lends itself to the expression of truths about that subject-matter. A particular metaphorical *utterance* is apt to the extent that (a) it is a move in an apt game, and (b) it makes impressive use of the resources that game provides. The reason it is so easy to have aptness without truth is that to make satisfying use of a game with lots of



expressive potential is one thing, to make veridical use of a game with arbitrary expressive potential is another.<sup>23</sup>

### CORRECTNESS IN NON-NORMAL MATHEMATICS

Back now to the main issue: what accounts for the feeling of a right and a wrong way of proceeding when it comes to mathematical theory-development? I want to say that a proposed new axiom  $A$  strikes us as correct roughly to the extent that a theory incorporating  $A$  seems to us to make for an *apter game*—a game that lends itself to the expression of more metaphorical truths—than a theory that omitted  $A$ , or incorporated its negation. To call  $A$  correct is to single it out as possessed of a great deal of ‘cognitive promise’.<sup>24</sup>

Take for instance the controversy early in the last century over the axiom of choice. One of the many considerations arguing *against* acceptance of the axiom is that it requires us to suppose that geometrical spheres decompose into parts that can be reassembled into multiple copies of themselves. (The Banach–Tarski paradox.) Physical spheres are not *like* that, so we imagine, hence the axiom of choice makes geometrical space an imperfect metaphor for physical space.

One of the many considerations arguing in *favor* of the axiom is that it blocks the possibility of sets  $X$  and  $Y$  neither of which is injectable into the other. This is crucial if injectability and the lack of it are to serve as metaphors for relative size. It is crucial that the statement about functions that ‘encodes’ the fact that there are not as many  $Y$ s as  $X$ s should be seen in the game to *entail* the statement ‘encoding’ the fact there are at least as many  $X$ s as  $Y$ s. This entailment would not go through if sets were not assumed to satisfy the axiom of choice.<sup>25</sup> Add to this that choice *also* mitigates the paradoxicality of the Banach–Tarski result, by opening our eyes to the possibility of regions too inconceivably complicated to be assigned a volume, and it is no surprise that choice is judged to make for an overall apter game. (This is hugely oversimplified, no doubt; but it illustrates the kind of consideration that I take to be relevant.)

Suppose we are working with a theory  $T$  and are trying to decide whether to extend it to  $T^* = T + A$ . An impression I do *not* want to leave is that  $T^*$ 's aptness is simply a matter of its expressive potential with regard to our original *naturalistic* subject matter: the world we really believe in, which, let's suppose, contains only concrete things.  $T^*$  may also be valued for the expressive assistance it provides in connection with the *mathematical* subject matter postulated by  $T$ —a subject-matter which we take to obtain in our role as players of the  $T$ -game. A new set-theoretic axiom may be valued for the light it sheds not on

<sup>23</sup> Calling a figurative description ‘wicked’ or ‘cruel’ can be a way of expressing appreciation on the score of aptness but reservations on the score of truth. See in this connection Moran (1989).

<sup>24</sup> Thanks to David Hills for this helpful phrase. <sup>25</sup> Thanks here to Hartry Field.



concreta but on mathematical objects already in play. So it is, for instance, with the axiom of projective determinacy and the sets of reals studied in descriptive set theory.

Our account of correctness has two parts. Sometimes a statement is correct because it is true according to an implicitly understood background story, such as Peano Arithmetic or ZFC. This is a relatively objective form of correctness. Sometimes, though, there is no well-enough understood background story and we must think of correctness another way. The second kind of correctness goes with a statement's 'cognitive promise', that is, its being of a type to figure in especially apt pretend games.

### OUR GOODMANIAN ANCESTORS

If mathematics is a myth, how did the myth arise? You got me. But it may be instructive to consider a meta-myth about how it might have arisen. My strategy here is borrowed from Wilfrid Sellars in *Empiricism and the Philosophy of Mind*. Sellars asks us to

Imagine a stage in pre-history in which humans are limited to what I shall call a Rylean language, a language of which the fundamental descriptive vocabulary speaks of public properties of public objects located in Space and enduring through Time. (Sellars, 1997: 91)

What resources would have to be added to the Rylean language of these talking animals in order that they might come to recognize each other and themselves as animals that *think*, *observe*, and have *feelings* and *sensations*? And, how could the addition of these resources be construed as reasonable? (Sellars, 1997: 92)

Let us go back to a similar stage of pre-history, but since it is the language's concrete (rather than public) orientation that interests us, let us think of it not as a Rylean language but a *Goodmanian* one. The idea is to tell a just-so story that has mathematical objects invented for good and sufficient reasons by the speakers of this Goodmanian language: henceforth *our Goodmanian ancestors*. None of it really happened, but our situation today is as if it had happened, and the memory of these events was then lost.<sup>26</sup>

#### First Day, Finite Numbers of Concreta.

Our ancestors, aka the Goodmanians, start out speaking a first-order language quantifying over concreta. They have a barter economy based on the trading of

<sup>26</sup> Earlier versions of this chapter had a fourteen-day melodrama involving functions on the reals, complex numbers, sets vs. classes, and more besides. It was ugly. Here I limit myself to cardinal numbers and sets.

precious stones. It is important that these trades be perceived as fair. To this end, numerical quantifiers are introduced:

$$\begin{aligned}\exists_0 x Fx &=_{df} \forall x (Fx \rightarrow x \neq x) \\ \exists_{n+1} x Fx &=_{df} \exists y (Fy \ \& \ \exists_n x (Fx \ \& \ x \neq y))\end{aligned}$$

From  $\exists_n \# \text{ruby}(x)$  and  $\exists_n \# \text{sapphire}(x)$ , they infer 'rubies-for-sapphires is a fair trade' (all gems are considered equally valuable). So far, though, they lack premises from which to infer 'rubies-for-sapphires is *not* a fair trade'. If they had infinite conjunction, the premise could be:

$$\sim(\exists_0 x Rx \ \& \ \exists_0 x Sx) \ \& \ \sim(\exists_1 x Rx \ \& \ \exists_1 x Sx) \ \& \ \text{etc.}$$

But their language is finite, so they take another tack. They decide to make as if there are non-concrete objects called 'numbers'. The point of numbers is to serve as measures of cardinality. Using  $*S*$  for 'it is to be supposed that  $S$ ', their first rule is:

- ⊙ (R1) If  $\exists_n x Fx$  then  $*n$  = the number of  $Fs*$ , and if  $\sim\exists_n x Gx$  then  $*n \neq$  the number of  $Gs*$ <sup>27</sup>

From  $(\#x)Rx \neq (\#x)Sx$ , they infer 'rubies-for-sapphires is not a fair trade'. ('The number of  $Fs$ ' will sometimes be written ' $(\#x)Fx$ ' or ' $\#(F)$ '.) Our ancestors do not believe in the new entities, but they pretend to for the access this gives them to a fact that would otherwise be inexpressible, viz., that there are (or are not) exactly as many rubies as sapphires.

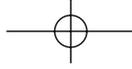
### Second Day, Finite Numbers of Finite Numbers.

Trading is not the only way to acquire gemstones; one can also inherit them, or dig them directly out of the ground. As a result some Goodmanians have more stones than others. A few hotheads clamor for an immediate redistribution of all stones so that everyone winds up with the same amount. Others prefer a more gradual approach in which, for example, there are five levels of ownership this year, three levels the next, and so on, until finally all are at the same level. The second group is at a disadvantage because their proposal is not yet expressible. Real objects can be counted using (R1), but not the pretend objects that (R1) posits as measures of cardinality. A second rule provides for the assignment of numbers to bunches of pretend objects:

$$(R2) \ * \text{If } \exists_n x Fx \text{ then } n = (\#x)Fx*, \text{ and } * \text{If } \sim\exists_n x Gx \text{ then } n \neq (\#x)Gx*$$

The gradualists can now put their proposal like this:  $* \text{every year should see a decline in the number of numbers } k \text{ such that someone has } k \text{ gemstones.}*$  The new rule also has consequences of a more theoretical nature, such as  $* \text{every}$

<sup>27</sup>  $F$  and  $G$  are predicates of concreta.



number is less than some other number.\* Suppose to the contrary that \*the largest number is 6.\* Then \*the numbers are 0, 1, 2, . . . , and 6.\* But \*0, 1, 2, . . . , and 6 are seven in number.\* So by (R2), \*there is a number 7\*.

**Third Day, Operations on Finite Numbers.**

Our ancestors seek a uniform distribution of gems, but find that this is not always so easy to arrange. Sometimes indeed the task is hopeless. Our ancestors know some sufficient conditions for ‘it’s hopeless’, such as ‘there are five gems and three people’, but would like to be able to characterize hopelessness in general. They can get part way there by stipulating that numbers can be added together:

$$(R3) \text{ *If } \sim \exists x (Fx \ \& \ Gx), \text{ then } \#(F) + \#(G) = \#(F \vee G)*.$$

Should there be two people, the situation is hopeless iff  $\sim \exists n \#(\text{gems}) = n + n^*$ . Should there be three people, the situation is hopeless iff  $\sim \exists n \#(\text{gems}) = ((n + n) + n)^*$ . A new rule:

$$(R4) \text{ *If } m = \#(G), \text{ then } \#(F) \times \#(G) = \#(F) + \dots + \#(F)^* \text{ (} m \text{ times).}$$

allows them to wrap these partial answers up into a single package. The situation is hopeless iff  $\sim \exists n \#(\text{gems}) = n \times \#(\text{people})^*$ .

**Fourth Day, Finite Sets of Concreta.**

Gems can be inherited from one’s parents, and also from their parents, and theirs. However our ancestors find themselves unable to answer in general the question, ‘from whom can I inherit gems?’ This is because they lack (the means to express) the concept of an ancestor. They decide to make as if there are finite sets of concreta:

$$(R5) \text{ For all } x_1, \dots, x_n, \text{ *there is a set } y \text{ such that for all } z, z \in y \text{ iff } z = x_1 \vee z = x_2 \vee \dots \vee z = x_n^*.$$
<sup>28</sup>

Ancestorhood can now be defined in the usual way. An ancestor of *b* is anyone who belongs to every set containing *b* and closed under the parenthood relation. Now our ancestors know (and can say) who to butter up at family gatherings: their ancestors.

**Fifth Day, Infinite Sets of Concreta.**

Gemstones are cut from veins of ruby and sapphire found underground. Due to the complex geometry of mineral deposits (and because miners are a quarrelsome

<sup>28</sup> *n* here is schematic.



lot), it often happens that two miners claim the same bit of stone. Our ancestors decide to systematize the conditions of gem discovery. This much is clear: Miner Jill has discovered any (previously undiscovered) quantity of sapphire all of which was noticed first by her. But how should other bits of sapphire be related to the bits that Jill is known to have discovered for Jill to count as discovering those other bits too? One idea is that they should *touch* the bits of sapphire that Jill is known to have discovered. But the notion of touching is not well understood, and it is occasionally even argued that touching is impossible, since any two atoms are some distance apart. Our ancestors decide to take the bull by the horns and work directly with sets of atoms. They stipulate that:

- (R6) If  $F$  is a predicate of concreta, then \*there is a set  $y$  such that for all  $z$ ,  
 $z \in y$  iff  $Fz^*$ ,

and then, concerned that not all sets of interest are the extensions of Goodmanian predicates, boot this up to:

- (R7) Whatever  $x_1, x_2, \dots$  might be, \*there is a set containing all and only  
 $x_1, x_2 \dots$  <sup>\*29</sup>  $\odot$

Next they offer some definitions. Two sets  $S$  and  $T$  of atoms *converge* iff given any two atoms  $x$  and  $y$ , some  $s$  and  $t$  in  $S$  and  $T$  respectively are closer to one another than  $x$  is to  $y$ .<sup>30</sup> A set  $U$  of atoms is *integral* iff it intersects every set of atoms converging on any of its non-empty subsets. A set  $V$  of atoms all of the same type—sapphire, say—is *inclusive*, qua set of sapphire atoms, iff  $V$  has as a subset every integral set of sapphire atoms on which it converges. The sought after principle: Jill can lay claim to the contents of the smallest inclusive set of sapphire atoms containing the bit she saw first.

### Sixth Day, Infinite Numbers of Concreta.

Numbers have not yet been assigned to infinite totalities, although infinite numbers promise the same sort of expressive advantage as finite ones. Our ancestors decide to start with infinite totalities of concreta, like the infinitely many descendants they envisage. Their first rule is:

- (R8) If  $\forall x \forall x' \forall y ((Rxy \ \& \ Rx'y) \rightarrow x = x')$  and  $\forall x (Fx \rightarrow \exists y (Gy \ \& \ Rxy))$ ,  
 then  $\#(F) \leq \#(G)^*$ . <sup>no bar on top</sup>

This is fine as far as it goes, but it does not go far enough, or cardinality relations will wind up depending on what relation symbols  $R$  the language happens to contain. Having run into a similar problem before, they know what to do.

<sup>29</sup> One might wonder how our ancestors acquired plural quantifiers, and whether they wouldn't have saved themselves a lot of trouble by acquiring them earlier.

<sup>30</sup> Crucially for this definition,  $x$  and  $y$  can be material or spatial atoms. Our ancestors hold that all point-sized spatial positions are occupied by points of space; material atoms cohabit with some of these but not all.



(R9) For each  $x$  and  $y$ , \*there is a unique ordered pair  $\langle x, y \rangle$ .\*

⊙ (R10) \*If  $p_1, p_2, \dots$  are ordered pairs of concreta, then there is a set containing all and only  $p_1, p_2, \dots$ .\* ⊙

A set that never pairs two right elements with the same left element is a function; if in addition it never pairs two left elements with the same right element, it is a 1–1 function; if in addition its domain is  $X$  and its range is a subset of  $Y$ , it is a 1–1 function from  $X$  into  $Y$ .

(R11) \*If a 1–1 function exists from  $\{x:Fx\}$  into  $\{x:Gx\}$ , then  $\#(F) \leq \#(G)$ .\*

How many infinite numbers this nets them depends on the size of the concrete universe. To obtain a *lot* of infinite numbers, however, our ancestors will need to start counting abstracta.

### Seventh Day, Infinite Sets (and Numbers) of Abstracta.

The next step is the one that courts paradox. (R7) allows for the unrestricted gathering together of concreta. (R10) allows for the unrestricted gathering together of a particular variety of abstracta. Now our ancestors take the plunge:

(R12) \*If  $x_1, x_2, \dots$  are sets, then there is a set  $\{x_1, x_2, \dots\}$ .\*

Assuming a set-theoretic treatment of ordered pairs, the sets introduced by (R12) already include the 1–1 functions used in the assignment of cardinality. Thus there is no need to reprise (R9); we can go straight to

⊙ (R13) \*If a 1–1 function exists from set  $S$  into set  $T$ , then  $\#(S\text{'s members}) \leq \#(T\text{'s members})$ .\* ⊙

(R12) will seem paradoxical to the extent that it seems to license the supposition of a universal set. It will seem to do that to that extent that ‘all the sets’ looks like it can go in for ‘ $x_1, x_2, \dots$ ’ in (R12)’s antecedent. ‘All the sets’ will look like an admissible substituent if the *de re* appearance of ‘ $x_1, x_2, \dots$ ’ is not taken seriously. But our ancestors take it *very* seriously. Entitlement to make as if there is a set whose members are  $x, y, z, \dots$  depends on *prior* entitlements to make as if there are each of  $x, y, z, \dots$ . Hence the sets whose supposition is licensed by (R12) are the well-founded sets.

### Much, Much Later, Forgetting.

These mathematical metaphors prove so useful that they are employed on a regular basis. As generation follows upon generation, the knowledge of how the mathematical enterprise had been launched begins to die out and is eventually lost altogether. People begin thinking of mathematical objects



as genuinely there. Some, ironically enough, take the theoretical indispensability of these objects as a *proof* that they are there—ironically, since it was that same indispensability that led to their being concocted in the first place.

#### WORKED EXAMPLE

An oddity of Quine's approach to mathematical ontology has been noted by Penelope Maddy (1997). Quine sees math as continuous with 'total science' both in its subject matter and in its methods. Aping a methodology he sees at work in physics and elsewhere, Quine maintains that in mathematics too, we should keep our ontology as small as practically possible. ~~Thus—~~

[I am prepared to] recognize indenumerable infinities only because they are forced on me by the simplest known systematizations of more welcome matters. Magnitudes in excess of such demands, e.g., beth-omega or inaccessible numbers, I look upon only as mathematical recreation and without ontological rights. Sets that are compatible with [Gödel's axiom of constructibility  $v = L$ ] afford a convenient cut-off. . . (1986: 400).

Quine even proposes that we opt for the 'minimal natural model' of ZFC, a model in which all sets are constructible *and* the tower of sets is chopped off at the earliest possible point. Such an approach is 'valued as inactivat[ing] the more gratuitous flights of higher set theory . . . ' (Quine, 1992: 95).

Valued by whom? one might ask. Not actual set theorists. To them, cardinals the size of beth-omega are not even slightly controversial. They are guaranteed by an axiom introduced already in the 1920s (Replacement) and accepted by everyone. Inaccessibles are far too low in the hierarchy of large cardinals to arouse any suspicion. As for Gödel's axiom of constructibility, it has been widely criticized—including by Gödel himself—as entirely too restrictive. Here is Moschovakis, in a passage quoted by Maddy:

The key argument against accepting  $v = L$  . . . is that the axiom of constructibility appears to restrict unduly the notion of an *arbitrary* set of integers (1980: 610).

Set-theorists have wanted to *avoid* axioms that would 'count sets out' just on grounds of arbitrariness. They have wanted, in fact, to run as far as possible in the other direction, seeking as fully packed a set-theoretic universe as the iterative conception of set permits. All this is reviewed in fascinating detail by Maddy; see especially her discussion of the rise and fall of Definabilism, first in analysis and then in the theory of sets.

If Quine's picture of set theory as something like abstract physics cannot make sense of the field's plenitudinarian tendencies, can any other picture do better? Well, clearly one is not going to be worried about multiplying entities if the entities are not assumed to really exist. But we can say more.

The likeliest approach if the set-theoretic universe is an intentional object more than a real one would be (A) to articulate the clearest intuitive conception possible, and then, (B) subject to that constraint, let all heck break loose.

Regarding (A), *some* sort of constraint is needed or the clarity of our intuitive vision will suffer. This is the justification usually offered for the axiom of foundation, which serves no real mathematical purpose—there is not a single theorem of mainstream mathematics that makes use of it—but just forces sets into the familiar and comprehensible tower structure. Without foundation there would be no possibility of ‘taking in’ the universe of sets in one intellectual glance.

Regarding (B), it helps to remember that sets ‘originally’ came in to improve our descriptions of non-sets. e.g., there are infinitely many  $Z$ s iff the set of  $Z$ s has a proper subset  $Y$  that maps onto it one–one, and uncountably many  $Z$ s iff it has an infinite proper subset  $Y$  that *cannot* be mapped onto it one–one. Since these notions of *infinitely* and *uncountably many* are topic neutral—the  $Z$ s do not have to meet a ‘niceness’ condition for it to make sense to ask how many of them there are—it would be counterproductive to have ‘niceness’ constraints on when the  $Z$ s are going to count as bundleable together into a set.<sup>31</sup> It would be still more counterproductive to impose ‘niceness’ constraints on the 1–1 functions; when it comes to infinitude, one way of pairing the  $Z$ s off 1–1 with just some of the  $Z$ s ~~seems~~<sup>is</sup> as good as another.  $\wedge$

So: if we think of sets as having been brought in to help us count concrete things, a restriction to ‘nice’ sets would have been unmotivated and counterproductive. It would not be surprising if the anything-goes attitude at work in those original applications were to reverberate upward to contexts where the topic is sets themselves. Just as we do not want to tie our hands unnecessarily in applying set-theoretic methods to the matter of whether there are uncountably many space–time points, we don’t want to tie our hands either in considering whether there are infinitely many natural numbers, or uncountably many sets of such numbers.

A case can be made, then, for (imagining there to be) a *plenitude* of sets of numbers; and a ‘full’ power set gathering all these sets together; and a plenitude of 1–1 functions from the power set to its proper subsets to ensure that if the power set isn’t countable, there will be a function on hand to witness the fact. Plenitude is topic-neutrality writ ontologically. The preference for a ‘full’ universe is thus unsurprising on the as-if conception of sets.

<sup>31</sup> Except to the extent that such constraints are needed to maintain consistency.

## REFERENCES

- Balaguer, Mark (1996). 'A Fictionalist Account of the Indispensable Applications of Mathematics.' *Philosophical Studies*, 83: 291–314.
- (2000). *Platonism and Anti-Platonism in Mathematics*. Oxford: Oxford University Press.
- (2001). 'A Theory of Mathematical Correctness and Mathematical Truth.' *Pacific Philosophical Quarterly*, 82: 87–114.
- Beaney, Michael (1997). *The Frege Reader*. Oxford: Basil Blackwell.
- Burgess, John P. and Gideon Rosen (1997). *A Subject With No Object*. Oxford: Clarendon Press.
- Field, Hartry (1980). *Science Without Numbers*. Princeton: Princeton University Press.
- (1989). *Realism, Mathematics and Modality*. Oxford: Basil Blackwell.
- Geach, Peter T. and Max Black (1960). *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell.
- Hahn, L. and P. Schilpp (eds.) (1986). *The Philosophy of W. V. Quine*. La Salle, IL: Open Court.
- Hills, David (1997). 'Aptness and Truth in Verbal Metaphor.' *Philosophical Topics*, 25: 117–153.
- Horgan, Terry (1984). 'Science Nominalized.' *Philosophy of Science*, 51: 529–49.
- Maddy, Penelope (1997). *Naturalism in Mathematics*. Oxford: Clarendon Press.
- Melia, Joseph (1995). 'On What There's Not.' *Analysis*, 55: 223–9.
- Moran, Richard (1989). 'Seeing and Believing: Metaphor, Image, and Force.' *Critical Inquiry*, 16: 87–112.
- Moschovakis, Y. (1980). *Descriptive Set Theory*. Amsterdam: North Holland.
- Putnam, Hilary (1971). *Philosophy of Logic*. New York: Harper & Row.
- Quine, Willard Van Orman (1951). 'Two Dogmas of Empiricism.' *Philosophical Review*, 60: 20–43. Reprinted in Quine (1961).
- (1961). *From a Logical Point of View*, second edition. New York: Harper & Row.
- (1986). 'Reply to Parsons,' in Hahn and Schilpp (1986).
- (1992). *Pursuit of Truth*, revised edition. Cambridge: Harvard University Press.
- Sellars, Wilfrid (1997). *Empiricism and the Philosophy of Mind*. Cambridge, MA: Harvard University Press.
- Steiner, Mark (1998). *The Applicability of Mathematics as a Philosophical Problem*. Cambridge, MA: Harvard University Press.
- Walton, Kendall L. (1990). *Mimesis and Make-Believe*. Cambridge, MA: Harvard University Press.
- (1993). 'Metaphor and Prop Oriented Make-Believe.' *European Journal of Philosophy*, 1.1: 39–57.
- Wigner, Eugene P. (1967). 'The Unreasonable Effectiveness of Mathematics in the Natural Sciences.' In *Symmetries and Reflections*. Bloomington, IN: Indiana University Press.

- Yablo, Stephen (1998). 'Does Ontology Rest on a Mistake?' *Proceedings of the Aristotelian Society*, Supplementary Volume, 72: 229–61, [Chapter 5 in this volume].
- (2002a). 'Go Figure: A Path through Fictionalism.' *Midwest Studies in Philosophy*, 25: 72–102 [Chapter 7 in this volume].
- (2002b). 'Abstract Objects: A Case Study.' *Philosophical Issues*, 12: 220–40 [Chapter 8 in this volume].
- 'Carving Content at the Joints' [Chapter 10 in this volume].



# 10

## Carving Content at the Joints

### 1. THE PROBLEM

Here is Frege in *Foundations of Arithmetic*, paragraph 64:

The judgment ‘Line  $a$  is parallel to line  $b$ ’, in symbols:  $a \parallel b$ , can be taken as an identity. If we do this, we obtain the concept of direction, and say: ‘The direction of line  $a$  is equal to the direction of line  $b$ ’. Thus we replace the symbol  $\parallel$  by the more generic symbol  $=$ , through removing what is specific in the content of the former and dividing it between  $a$  and  $b$ . We carve up the content in a way different from the original way, and this yields us a new concept (Frege 1997, 110–11).

Something important is going on in this passage. But at the same time it borders on incoherent. For Frege is saying at least the following:

- (1) ‘ $\text{dir}(a) = \text{dir}(b)$ ’ has the same content as ‘ $a \parallel b$ ’
- (2) reflecting on that can lead one to the concept of direction.

Why doesn’t (2) contradict (1)? (2) has a neophyte acquiring the concept of direction—and so presumably a grasp of the content of ‘ $\text{dir}(a) = \text{dir}(b)$ ’—by reflecting on a certain content–identity. But then it is hard to see how the postulated content–identity can really obtain; Leibniz’s Law would seem to forbid it. If one grasps content  $X$  at a certain time, and content  $X =$  content  $Y$ , then one grasps content  $Y$  at that time. The neophyte grasped the content of ‘ $a \parallel b$ ’ before encountering (1), so if that content is also the content of ‘ $\text{dir}(a) = \text{dir}(b)$ ’, she must have grasped the content of ‘ $\text{dir}(a) = \text{dir}(b)$ ’ before encountering (1) as well.

I know of only one good way of getting around this. The neophyte *did* grasp the content of ‘ $\text{dir}(a) = \text{dir}(b)$ ’ before encountering (1); she just failed to know it *as the content of an identity-sentence*. She doesn’t know it as the content of an identity-sentence until she acquires the concept of direction: perhaps knowing it that way *is* acquiring the concept of direction.

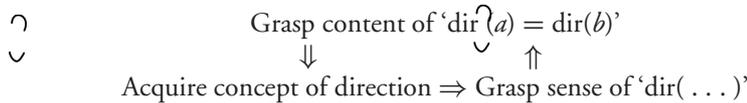
This paper was written for the Arché conference “Does Mathematics Require a Foundation?” (August 2002). Thanks to Bob Hale, Crispin Wright, Kit Fine, Neil Tennant, John Burgess, Gideon Rosen and Matti Eklund for very useful comments.



2. SENSE

What should content be, for this way around the problem to work? One natural hypothesis is that *content is sense*; and Frege certainly says things that suggest this. But the suggestion is problematic, if we take Frege at his word that the sense of part of a sentence is part of the sense of the sentence.

Remember, the neophyte has to grasp the content of ‘ $\text{dir}(a) = \text{dir}(b)$ ’ before acquiring the concept of direction. So if content is sense, she must be able to grasp the *sense* of ‘ $\text{dir}(a) = \text{dir}(b)$ ’ before acquiring the concept of direction. If she lacks the concept of direction, though, how is she supposed to grasp the sense of direction-terms? And if she does not grasp the sense of direction-terms, how is she supposed to grasp the sense of ‘ $\text{dir}(a) = \text{dir}(b)$ ’? The problem is that each of the indicated achievements presupposes the one “before” it:



Frege’s strategy does not appear to work, then, if content is sense. What else could it be?

The downward-facing arrow is compulsory, for the passage clearly states that the concept of direction is acquired by grasping the content of the direction-sentence. The left-to-right arrow is compulsory too, for grasping the sense of ‘direction-of’ is appreciating that it expresses the relevant concept. The upward-facing arrow is not forced on us, though. Why shouldn’t grasping the content of a sentence leave one still undecided about its sense?

The obvious way to arrange for that is to make content *coarser-grained* than sense, though presumably still finer-grained than reference. Then everything hangs together just right:

- (1) ‘ $\text{dir}(a) = \text{dir}(b)$ ’ shares *something* with ‘ $a \parallel b$ ,’ but
- (2) the shared something is content, not sense;
- (3) the shared content can be carved in two ways,
- (4) corresponding to the sentences’ two senses;
- (5) we start out knowing one carving, then learn the other;
- (6) the directional carving teaches us the concept of direction.

This has got to be the way to go. But every step raises questions, the main ones being (a) what is content and (b) what are carvings? I want to sketch a conception of carving that deals with some of these questions in a way broadly congenial to the Fregean platonist program in the foundations of mathematics.

## 3. SKETCH AND MOTIVATION

I can't claim *too* much of a Fregean basis for the proposal to follow. I would, however, like to relate the proposal to what are popularly regarded as some Fregean concerns and themes.

One such theme is that semantic theory begins with sentences and their truth values. Words and other subsentential expressions have their semantic values too; but these are constrained mainly by the requirement that they predict the expression's contribution to truth-value. Semantic values are whatever they have to be to deliver the right sentence-level results. One gets from the one to the other by a kind of abduction or inference to the best explanation.

A second theme concerns logical structure or form. There is no way of abducting semantic values that doesn't take a stand on the sentence's logical form. Without the form, one doesn't know what types of semantic value are needed (object, truth-function, *m*-place concept, etc.); and one doesn't know by what mode of combination they are supposed to deliver the truth-value. A given bunch of semantic values—say, Brutus, Caesar, and the relation of killing—might yield various truth-values depending on how they are combined.

Now, given what was said about semantic *values* being slaves to truth-value, it might seem that the *structures* the values are slotted into should likewise be whatever best serves the needs of truth-value prediction. A certain type of sentence may *look* atomic (or whatever), as "The dodo bird still exists" looks atomic. But if it proves hard to predict truth-values on that hypothesis, we may decide that its real, underlying, structure, is something quite different.

This leads directly to a third theme. I said you might *expect* Frege to make structure a slave to truth-value prediction. But you'd be wrong, or anyway not entirely right. He certainly does want to impute structures that generate the appropriate truth-values. But the enterprise is subject to a heavy constraint imposed by the actual words employed and how they are ordered.

This is just the familiar point that Frege tries very hard—much harder than Russell, to make the obvious comparison—not to run roughshod over grammatical appearances. He is concerned to understand the sentence as he finds it. Semantic structures ought if possible <sup>to</sup> parallel sentential structure, even if it takes lots of theoretical work to find a parallel semantic structure that does the job. Call that Frege's *empiricism about structure*.

To review, our three main players are *truth-value*, *semantic structure*, and *semantic value*. Of these three, the *first* is treated as given, and the *second* and *third* are reached abductively, subject to the constraint that semantic structures should not be imputed that run roughshod over the grammatical appearances.

This idea of respecting grammatical appearances is fair enough in itself, but it creates a tension in Frege's system. What if we spot an alternative, not entirely

parallel, semantic structure, that works better somehow than the structure we were initially inclined to impute? (I propose to be vague for the moment about what *better* might mean here. The alternative structure predicts truth-values more efficiently, perhaps, or in a way that illuminates entailment relations with other sentences, or . . .).

If we are Frege, we cannot make like Russell and adopt this alternative structure despite its grammatical implausibility. But we also don't want to just wave the alternative structure goodbye. Because while, on the one hand, we find fault with it for not running parallel enough to the sentence's apparent grammatical form, at a deeper level we may feel that it is *the sentence's fault* for not running parallel enough to this excellent structure. That we employ the sentence we do, and not one with the alternative structure, is *too bad* in a way. There would be *advantages* to using a sentence of the second sort rather than the one we do use.

What is the Fregean to do when this sort of feeling comes over him? I want to suggest that this is where the need for alternative carvings arises. Frege's *empiricism* (about structure) tells him that S's truth value is generated in one way. His *rationalism* tells him that it might better have been generated another way, the way S\*'s truth value is generated. Content (re)carving gives Frege an outlet for the second feeling that lets him stay true to the first. The conflict is resolved by noting that S's content can also be carved the S\* way, followed perhaps by a recommendation that S\* be used instead of S in situations where, as Quine later puts it, greater theoretical profundity is professed.

Consider in this light the discussion in paragraph 57 of adjectival versus singular statements of cardinality. "Since what concerns us here is to define a concept of number that is useful for science, we should not be put off by the attributive form in which number also appears in our every day use of language. This can always be avoided. . . the proposition 'Jupiter has four moons' can be converted into 'The number of Jupiter's moons is four', [where] 'is' has the sense of 'is equal to', 'is the same as'" (Frege 1997, 106–7). Yes, we do use the attributive form (that's the empiricism talking), but it is possible, and Frege suggests more perspicuous, to convey the information with an identity statement. This occurs shortly before paragraph 64 on content carving and foreshadows his principal application of the notion.

The *motivation* for carving is normative, or ameliorative: it would be *better* if the job had been given to S\* instead of S. This not to say that the notion of carving is itself normative. Rather, the normative claim has a factual presupposition, and carving speaks to that presupposition. S\* can't do the job better than S unless there is a thing they both do, each in its own way. *Content* is Frege's word for the thing they both do; *carving* is his word for their different ways of doing it. All of that is perfectly factual. I'll be suggesting, however, that the factual notion comes into clearer focus if we remember the ameliorative motivation.

## 4. ROADS TO CONTENT

What is this thing *content*, the expressing of which is the job  $S$  and  $S^*$  have in common? One thing we know is that it lies somewhere between truth value and sense. I want to sketch a route up to content from truth-value and a second route down from sense.

## Up from Truth-Value

A semantic theory is *materially adequate* if it assigns semantic values and semantic structures that together yield sentences' actual truth-values. It may seem that the fundamental semantic project for a given language is to come up with a materially adequate interpretation of the language. But on reflection that is not quite right.

A theory could be materially adequate for the wrong reasons. It could be a matter of *luck* that it succeeds in making all the truths come out true and the falsehoods false; the theory is able to copy only because the facts take a particularly tractable form. An accidentally adequate theory would not have been acceptable to Frege. It is hard to imagine him saying, "Let's hope it snows tonight, for if not, then truth-values will be distributed in a way I am powerless to predict, given the structures I have assigned."

Imagine that we mistakenly treated "~~someone~~" as a name. This would make it very hard to assign semantic values to basic expressions so as to generate the correct truth-values. For let  $F$  be a predicate that is true of some things but not of everything. "~~Someone~~ is  $F$ " and "~~someone~~ is not- $F$ " ought to both come out true. Whatever value  $s$  we assign to the "name," though, this is not the result we obtain; we can make  $\sim F_s$  true only by making  $F_s$  false.

Well, but we might get lucky. It might happen that every predicate of the language was satisfied either by everything or by nothing. Then we wouldn't *want*  $F_s$  and its negation to both come out true. It wouldn't matter what we assigned to  $s$ ; if  $F_s$  was true (false) on one assignment, it would be true (false) on all of them. Our theory would escape refutation by pure dumb luck.

A good theory should not depend for its material adequacy on lucky accidents. That is essentially to say that one needs a *policy* of semantic value assignment that, no matter how the world turns out, assigns values of a type that, plugged into the relevant structures, takes one to the correct destination, true or false. The policy should say for each expression  $E$ , and every situation  $W$  in which we might find ourselves, that

$E$ 's semantic value on the  $W$ -hypothesis is so and so.

What determines if the policy is a good one? Well, sentences have to come out with the right truth *profiles*. That is, we first ask what truth-value a sentence

*S* deserves on the hypothesis that we are in situation *W*. Then we ask what truth-value the sentence receives if basic expressions are assigned values according to the policy. A successful assignment policy has these always coming out the same.

So far, so good, but where do contents come in? They are already in. An *assignment policy* is no different from a series of functions—one per expression *E*—taking circumstances *W* to *E*'s semantic value *SV* in *W*. A *truth profile* for sentence *S* is no different from a function taking circumstances *W* to *S*'s truth-value in *W*. Value profiles are my candidates for the role of *E*'s content and truth profiles are my candidate for the content of sentence *S*.

This identification having been made, the project of assigning semantic values to basic expressions that yield the expected truth-values *non-accidentally* is the same as the project of assigning *contents* to basic expressions that yield the expected *contents* for sentences. To whatever extent the first project is Frege's, the second is too, albeit formulated in a way he might not recognize or appreciate.

### Down from Sense

The basic work of a sentence is to be true or false. Of course, the sentence is true or false because of its sense, or the thought expressed. But there is liable to be more to the thought than is needed to determine its truth value (examples in a moment). One might want to abstract away from this excess and limit attention to those aspects of the thought <sup>that are</sup> potentially relevant to truth.

This is not so different from what Frege himself does when he abstracts away from tone and color and from the “hints” given by words like “still” and “but.” According to Frege, “Alfred has still not arrived” *hints* that Alfred is expected, but this has no bearing on truth—Alfred's turning out not to be expected would not make the sentence false—so Frege leaves these aspects of meaning out of the thought. The idea here is to continue Frege's project of purging whatever is truth-irrelevant and focussing on what is left. *One* natural stopping point is the thought, but it is possible to go further.

Take ‘Today is sunny’ (uttered on August 16, 2002) and ‘August 16, 2002 is sunny’ (this is an example of Michael Beaney's). What do they share? Not sense, because accepting one does not rationally commit me to accepting the other. More than truth-value, though, because truth-value ignores that the thoughts stand or fall together. Beaney remarks that

An obvious candidate is Frege's early notion of conceptual content, which, if a metaphysical gloss could be put on the notion, might be best characterized as referring to [Umstände or] ‘circumstances’ (Frege 1997, 34–5; see 52–3, *Begriffsschrift* 2, for Umstände).

The sameness of content here seems well captured by our idea of senses whose differences are guaranteed in advance not to make a difference to truth-value.

The thought expressed by ‘Today is sunny’ uttered on August 16 cannot differ in truth-value from the the one expressed by ‘August 16 is sunny.’ Likewise the thoughts expressed by ‘Gustav Lauben is thinking,’ spoken by Gustav Lauben, and ‘I am thinking,’ written by Lauben at the same time.

Of course these indexicality-based examples can hardly serve as a model of the relation between “lines  $a$  and  $b$  are parallel” and “the direction of  $a$  is identical to the direction of  $b$ .” An example of Davidson’s comes a bit closer (Davidson 1979). Suppose that everything has exactly one shadow, and vice versa. Associated with each name  $a$  there is a name  $a^\#$  that stands for the  $a$ -object’s shadow. To each predicate  $P$  corresponds a predicate  $P^\#$  that is true of a shadow iff the object casting the shadow is  $P$ . If  $S$  is an atomic sentence  $Pa$ , let  $S^\#$  be  $P^\#a^\#$ . Clearly  $S$  and  $S^\#$  differ in sense; only one involves the concept of shadow. But the difference is of no possible relevance to truth-value. However matters might stand,  $S$  is true iff  $S^\#$  is. They therefore agree in content, if content is the truth-relevant aspect of sense. There are other ways of pulling the same basic trick, such as Quinean proxy functions.

A third example comes from the “slingshot” argument—an argument which has been seen as clarifying and/or consolidating Frege’s reasons for rejecting a level of significance between sense and truth-value, such as content is supposed to be. How does the argument go? Let  $S$  and  $T$  be both true or both false. Then it seems that the following ought to be a significance-preserving sequence:

1.  $S$
2.  $0 =$  the number which is 0 if  $S$  and 1 if not- $S$ .
3.  $0 =$  the number which is 0 if  $T$  and 1 if not- $T$ .
4.  $T$

1. and 2. differ in sense, because only one involves the concept of a number. But this difference is (so one might claim) of no possible relevance to truth-value, hence the sentences have the same content. The same applies to 3. and 4. If 2. and 3. share a content, we are sunk, because all truths will wind up with the same content.

If content is the aspect of sense bearing on whether a sentence is true or false, the question of whether 2. and 3. share a content boils down to this: is the substitution of  $T$  for  $S$  potentially relevant to truth-value? Given our stipulation that  $S$  and  $T$  are both true or both false, it may seem the answer is NO. But this is to confuse *epistemic* relevance—what might *for all we know* change the truth value—with the intended semantic notion—what is in a *position* to change the truth value, even if we happen to know that in this case it does not. The slingshot argument gives no reason for rejecting content as we are beginning to conceive it here.

## 5. WHAT CONTENT IS

Two routes to content have been described. The first was a route up from truth-value, where the upward pressure was exerted by the *non-accidental* character of a theory's success at predicting truth-values. The second was a route down to content from sense, where the downward pressure came from our desire to bleach out aspects of sense with no possible bearing on truth.

One hopes, of course, that the two routes will converge on the same point. And this appears to be the situation; for to say of the thoughts expressed by S and T that their differences are of no possible relevance to truth value is to say that they are both true, or both false, *no matter what*. This gives us a first rough definition of what is involved in sharing a content.

(CONTENT—intuitive) S and T share a content iff the thoughts they express, although perhaps different, differ in a way that makes no (possible) difference to truth.

One can imagine ways of elaborating this. One could ask, e.g., that it be *knowable a priori* that S and T have the same truth value no matter what. And one could ask that it be a priori knowable *independently of any intelligence one might possess about what the sentences' truth-values actually are*.<sup>1</sup> But these more elaborate approaches, although they get us a same-content relation on sentences, do not get us all the way to contents considered as entities in their own right. (The most we could hope for is equivalence classes of sentences.) Because the present approach calls for *contents*, we cannot afford to be so fancy. In this paper, “same truth-value no matter what” means: true in the same cases.

This might be thought un-Fregean for the following reason. Cases sound a lot like worlds; and we are told that “Frege has no notion of metaphysically possible worlds distinct from this world,” and indeed rejects “metaphysical modality” altogether (Levine 1996, 168).

But to say that he had no use for *metaphysical* modality is not to say that Frege rejects modality altogether. He accepts a priority and an epistemicized version of analyticity (uninformativeness): and more to the present point, the notion of sense is implicitly modal. Frege explains sense in more than one way, of course, but there is a clear modal element in sense qua *mode of determination* (Frege 1997, 22–23).

People sometimes object to the mode of determination account that sense cannot determine reference *all by itself*. If it did, then merely understanding a

<sup>1</sup> This is to allow for differences in content between sentences both of which are knowable a priori, e.g., “Sisters are siblings” and Fermat's Last Theorem.

sentence would put you in a position to know its truth value. This is to read “determines the referent” as “leaves no room for other factors, such as the way of the world.” A more plausible reading is “exhibits the referent as a function of those other factors.” I don’t see what it can mean to say that the sense of “the Evening Star” determines its reference if not that the reference is one thing if, say, the body visible in the evening is Mars, another if it is Venus. Similarly, what could it mean to say that the thought expressed by a sentence determines its truth value, if not that the truth-value is one thing if the world is this way, another thing if not?

That Frege accepts some sort of modality—call it “conceptual” modality—might seem no help, because possible worlds are suited to the explanation only of metaphysical modality. But this is in fact controversial. Some philosophers maintain that there are two quite different ways of associating sentences with worlds, one of which lines up with conceptual necessity more than metaphysical.

How is this supposed to go? When Kripke talks about the worlds in which S, he means the *w*’s that answer to the description that S gives of our world. I will call that *satisfaction*. A world *satisfies* S iff it would have been that S, had *w* obtained. When Fregeans talk, to the extent that they can be induced to talk, about worlds in which S, they mean the *w*’s such that if this turns out to be *w*, then S. I will call that *verification*. A world *verifies* S iff S holds on the supposition that *w* really does obtain.

So, to go with the usual example, consider a world *w* where Venus appears in the evening but the planet appearing in the morning is Mars. This world doesn’t *satisfy* “Hesperus isn’t Phosphorus,” because it is not true that if certain appearances had had different causes, Hesperus, that is Venus, would have been distinct from Phosphorus, that is, Venus. But the world I mentioned *does* verify “Hesperus isn’t Phosphorus,” for if astronomers have *in fact* misidentified the morning-visible planet—it’s really Mars—then Hesperus *isn’t* Phosphorus.

Now clearly the mode of evaluation relevant to sense is verification; to say that the thought determines a truth value is to say that whether it is actually true or actually false depends on what (actually) happens. But then, given that content is a coarsening of sense, the mode of evaluation bearing on content ought to be verification too.<sup>2</sup> I find it is easier to keep the *verification* aspect clearly in mind if we speak not of worlds but *cases*. (“Have you heard? The morning-visible planet turns out not to be Mars.” “In that *case*, Hesperus is not Phosphorus.”) So the proposal is that

(CONTENT-official) S and T share a content iff they are true in the same cases. Contents are sets of cases.

<sup>2</sup> “But does the proposition ‘The Earth has two poles’ mean the same as ‘The North Pole is different from the South’? Obviously not. The second proposition could be true without the first being so, and vice versa” (Grundlagen 44).

Once again, “Hesperus = Phosphorus” has a contingent content, since it is not true in all cases. This is what you would hope and expect if the modality involved is non-metaphysical, since it is only in a metaphysical sense that Hesperus could not have failed to be Phosphorus. A sentence that is true in all cases is Evans’s “Julius invented the zip, if any one person did.” This is necessary not in a metaphysical sense—it could have been Julius’s mother that invented the zip—but conceptually—it could not turn out that the inventor wasn’t Julius.

## 6. CONFLATION

So we don’t need to worry that contents explained as sets of worlds are objectionably *modal*. Some related worries cannot be laid to rest so easily.

One is that there are not enough contents to go around, with the result that sentences that ought intuitively to be assigned *different* contents will be forced to share. I will call this the Conflation Problem.<sup>3</sup> Consider “Julius (if he existed) invented the zip,” “Sisters are siblings,” and “There is no largest prime number.” These are true, let’s suppose, in all cases, hence in the same cases. Yet one doesn’t feel that “Julius invented the zip” recarves the content of the Prime Number Theorem.

Bob Hale suggests an interesting answer to this objection, though he doesn’t accept the answer himself. The objection would succeed, he says,

if the claim were that two sentences having the same content is not only necessary but also sufficient for one to be properly viewed as recarving the content of the other. But the defender of the Fregean account has no need to make so strong a claim: he can claim that coincidence in truth-conditions . . . suffices as far as the requirement of identity of content goes, but point out that this does not preclude the imposition of further conditions on the sentences involved (Hale 1997, 95).

Michael Potter and Timothy Smiley find this baffling: “Hale is suggesting, twice over, that two sentences can have the very same content but not count as recarvings of that content. This seems to us incomprehensible” (Potter and Smiley 2001, 328).

Once we draw a certain distinction, however, Hale’s position is no longer at odds with that of Potter and Smiley. The distinction is between *S*’s *tolerating* *S*\*-style recarving, and its *inviting* *S*\*-style recarving. (Smiley Potter and are right about the first; if two sentences share a content, then each tolerates the recarving of its content provided by the other. But Hale’s remarks can be read as directed at

<sup>3</sup> One version of this *has* been answered, viz., that Kripkean a posteriori necessities will share a content—the necessary content—despite conveying very different empirical information. Our answer to this is that Kripkean a posteriori necessities aren’t true in all cases, or in the same cases. There are plenty of things we could learn that would lead us to say, “if that is really the case, then this lectern is made of ice.”

the second, and then they seem entirely sensible. *S invites* an  $S^*$ -style recarving of their shared content only if the new carving improves somehow on the original; and most ways of recarving a content are just different, not better. There will be more on this after we consider the Proliferation Problem.

## 7. PROLIFERATION

There is a way of putting Proliferation that makes it sound just like Conflation. Conflation occurs if

too many thoughts are recarvings of one content.

Proliferation occurs if

one content admits recarving into too many thoughts.

The difference is a matter of emphasis. Conflation puts it on *one content*. Each of the carvings may be in its own way legitimate; but they shouldn't all be of the same content. Proliferation puts the emphasis on *too many thoughts*. There is no problem about these thoughts' carving the same content, supposing them to be otherwise admissible; but lots of them *aren't* otherwise admissible. This worry arises in a particularly sharp form on the conception of content proposed above.

Suppose that contents are sets of cases, and that *S* and *T* share a content *C*. What is it for them to carve *C* differently? To have a specific example, *S* might be the conjunction of  $A_1$  and  $B_1$ , and *T* the disjunction of  $A_2$  and  $B_2$ .  $A_1 \& B_1$  carves its content conjunctively by exhibiting that content as arrived at by taking the intersection of two other contents, those of  $A_1$  and  $B_1$  respectively.  $A_2 \vee B_2$  carves that same content disjunctively, because it represents it as obtained by taking the union of  $A_2$ 's content with  $B_2$ 's. *S* and *T* carve the content differently because they exhibit it as constructed along different lines.

Carvings on this view are semantic etiologies or constructional histories. I will usually confine myself to immediate history, though ideally one would want to reach further back. A complete constructional history would be a structure tree of the kind found in categorial grammar textbooks. I will be worrying only about the top of that tree.

Now clearly, there is no backward road from a set to its history. Sets can be constructed in millions of ways, limited only by the ingenuity of the constructor. It helps a little to restrict the modes of construction to intersection, complement, and other functions expressed by logical devices present in natural language. But it doesn't help very much. Just as every number is a sum, difference, product, and so on, many times over, every set of cases is a union, intersection, complement, and so on, many times over.

Someone might say, what's wrong with that? Let a hundred flowers bloom. Maybe the resistance is just aesthetic and can be overcome.

But it is not just aesthetic. The resistance has to do with the role content carving is supposed to play in the introduction (or revelation) of objects. Initially, one is suspicious of certain objects and reluctant to accept them as real. Then it is pointed out that they are <sup>discernible</sup> ~~quantified~~ over in recarvings of contents one already accepts. This is supposed to be reassuring. The objects were already there lying in wait; they spring into view as soon as we set our logical microscope to the correct power.

This does sound reassuring. But not if it turns out that there are no controls on the operation—that objects of practically whatever type you like can be discerned in contents of practically whatever type you like. And this is a very real danger, if all it takes to discern a type of object in a content is to work that type of object into the calculation by which the content is obtained.

Do we really need Hume's Principle to exhibit arithmetic as already implicitly there in the contents of sentences we accept? If incorporating numbers into a constructional history is enough, then it can be done a lot more easily. Start with any sentence you like, say, "I reckon upon a speedy dissolution."<sup>4</sup> One route to its content is to look for the cases where Hume reckoned on a speedy dissolution. Another is to look for the worlds where he reckoned on a speedy dissolution and Peano's Axioms hold. You get the same worlds either way. Surely, though, we *don't* want to say that numbers can be discerned in the content of "I reckon upon a speedy dissolution."

It might be held that numbers *can* harmlessly be worked into any old content once we've got them—once we've obtained a guarantee of their existence. But to obtain that guarantee, you need a recarving with the right sort of epistemological backing. This is what Hume's Principle was supposed to provide, and abstraction principles more generally. To the extent that these can be regarded as merely concept-introducing—as teaching us what a direction or number is supposed to *be*—they seem well positioned to give us the required guarantee.

But problems also arise with recarvings backed by abstraction principles. I am not thinking here of the Bad Company Objection, which points to superficially similar principles—Frege's Basic Law (V) or Boolos's Parity Principle—that threaten contradiction. Suppose that inconsistency-threatening principles can be cordoned off somehow. We are still left with what is after all a more common problem with Company: that of being Uninvited, Unhelpful, and Unwelcome.<sup>5</sup> The Uninvited Company Objection, as I will call it, goes like this: Principles superficially similar to Hume's can be used to introduce perfectly consistent objects which have, however, nothing to recommend them. If the case for numbers is no better than for some of these other objects (see below), then it is hard to see why anyone but the extreme

<sup>4</sup> This was Hume's comment on learning that his condition was "mortal and incurable" ("My Own Life").

<sup>5</sup> The paper's working title was "Visiting relatives can be boring."

ontological maximalist should take numbers seriously.<sup>6</sup> Heck puts the problem as follows:<sup>7</sup>

Let  $xQy$  be an equivalence relation, chosen completely at random: It might, for example, have as one of its equivalence classes the set containing each of my shoes, my daughter Isobel, the blackboard in Emerson 104, and some other things. We can now introduce names purporting to stand for objects of a certain sort, call them *duds*, just as we introduced names of shapes and [directions]:

$$\text{dud}(a) = \text{dud}(b) \text{ iff } aQb.$$

But are we really to believe that there are such abstract objects as duds? To take a less random case, consider the equivalence relation: person  $x$  has the same parents as person  $y$ . In terms of this relation, we can introduce names of what I shall call *daps*. But are there such abstract objects as daps? In so far as one has an intuition that there are no such objects . . . there ought to be a corresponding doubt whether the neo-Fregean explanation of names of [shapes and directions] explains *these* names in such a way that they must denote abstract objects . . . Nothing in the neo-Fregean story distinguishes these cases.

Heck calls this the Proliferation Problem. Hale in a very subtle discussion gives some less artificial examples:<sup>8</sup>

can we really believe that our world contains, alongside our PM, that lady's *whereabouts*, and that in addition to Smith's murderer, there is another object, *his identity*, and, besides the claimant, his or her *marital status*? (Hale 1988, 22).

It would be silly to suppose that numbers had no better claim on our attention than this lot; one doesn't want to throw the baby out with the bathwater. But, and this is the worry, one doesn't want to take the bathwater in with the baby either.

## 8. BAD CARVING AND MATERIAL FALSITY

I spoke earlier of Frege's *rationalism*. Tyler Burge shows (Burge 1998, 1984, 1992) that this rationalism runs very deep. Frege believes in a *natural order of thoughts* to which human cogitation is naturally drawn. These thoughts are grasped obscurely to begin with, but more and more clearly as inquiry progresses. When Cauchy and Weierstrass gave their epsilon-delta definition of a limit, they did not replace one lot of calculus thoughts with another, so much as clarify the thoughts that people had already had.

Frege also holds that there are *objective laws of truth* charting the relations between thoughts. And he arguably also holds that the more a thought's entailment relations are subsumable under laws of truth, the better the thought is. The reason, or one reason, that epsilon-delta thoughts are so good is ↗

<sup>6</sup> Eklund (2006)

<sup>7</sup> Heck (2000)

<sup>8</sup> Hale (1988), p. 22.

that they turn what would otherwise be analytic entailments (trading on special features of limits, or of infinitesimals) into maximally general logical entailments.

Now I want to sketch a different rationalist theme whose role in Frege's thinking has not been much discussed.

Descartes gave our ideas two kinds of representational task. First is the task of *standing for whatever it is that they stand for*. Second is the task of *giving a non-misleading impression of that something*. Success at the first task by no means ensures success at the second; indeed, failure at the second task presupposes success at the first. An idea has to reference something before it can count as giving a wrong impression of that something.

Our idea of pain does a fine job, he thinks, where referring to pain is concerned. But it gives a confused or misleading impression of what pain is. Our ideas of heat and cold leave it unclear whether heat is the "real and positive" partner, and cold merely the absence of heat, or the other way around. Redness looks to be an intrinsic property of the red object, but it is really something else, perhaps a disposition to cause reddish sensations. Locke complains about secondary quality ideas quite generally that they fail to resemble the properties they are ideas of. Ideas that misrepresent, in the sense of giving a false or misleading impression of, their objects are called (by Descartes) *materially false*.

Now, Frege is not very interested in ideas. His preferred representational vehicle is the sentence. But a similar distinction can perhaps be made with respect to them. A sentence might succeed in expressing a certain content, while giving a wrong impression of that content.

The analogy may seem strained, to begin with. What sort of impression do sentences give, after all, of their contents? Well, a sentence containing a name of Socrates might give the impression of expressing a content in which Socrates figures. A sentence with a certain kind of logical structure might give the impression of expressing a content that is structured the same way.

Here the analogy breaks down, one might think. How pain and color can match up with our ideas of them is clear enough. (The idea of redness presents it as intrinsic, and this impression is either right or wrong depending on what redness is really like.) But it is not initially clear how contents—sets of cases—could have things (Socrates) figuring in them, or possess a certain logical structure.

The proper name case is easier. "Superman isn't real" gives the impression of being true in cases where there is an individual *Superman* with the property of not being real. X figuring in a content would be X existing in the cases that make up that content.

But how can a sentence be misleading about the logical structure of its content? Wouldn't that require contents to be structured, and aren't sets precisely unstructured?

They are certainly not explicitly structured. It could be, however, that sets of cases "lend themselves" to a certain style of decomposition, as the set of muskrats

lends  
 and bees ~~lending~~ itself to decomposition into the set of muskrats and the set of bees.

An analysis of this “lends itself” talk is suggested by David Lewis. A set of particulars is disjunctive, he says, if it is the union of two sets each of which is much more natural than it is (Lewis and Langton 2001). The same should hold for sets of cases or (as I’ll now say) worlds. The worlds with lots of neutrinos or lots of dragons in them decompose naturally into the lots-of-neutrino worlds and the lots-of-dragons ones. This, Lewis says, is because (i) the first set is the union of the second and the third, and (ii) the first set is much less natural than the second and third are. A similar analysis suggests itself of negative contents and perhaps also conjunctive ones.

Even if contents are sets, then, it is not out of the question that sentence S should give a “materially false” account of its content. S performs wonderfully at its primary task of expressing the relevant content. But it gives a misleading impression of that content, because the content is disjunctive and S is of the form A&B.

Suppose we revisit the Conflation and Proliferation problems with these notions in mind. Regarding Conflation, Hale suggests that sentences with the same content could nevertheless fail to count as alternative carvings of that content. Potter and Smiley find this incomprehensible. I said that both sides can be right, once we distinguish S *tolerating* recarving by S\*, and its *inviting* that kind of recarving.

One key element in S *inviting* recarving by S\* is that S\* does better at what we have called its secondary representational task; it exhibits the relevant content as put together in a way that is truer to that content’s internal nature.<sup>9</sup>

Our answer to the Conflation problem is this. When sentences share a content, each is indeed amenable to recarving in the style of the other. But these recarvings will normally be uninvited and unilluminating. There is nothing objectionable about a content’s *tolerating* lots of recarvings, so long as it doesn’t invite all of them.

Something similar applies to Proliferation. Unwanted objects may be *discernible* in lots of contents—but that will be because the carving was uninvited. Our policy should be to recognize only the entities that cry out to be recognized, because their contents lend themselves to quantificational carving. Of course, I haven’t yet said how the quantificational case is supposed to go, and what I do say might be found unconvincing. But this should not distract from the key distinction: objects revealed when a content *cries out* for quantificational recarving vs. objects “revealed” when a content *tolerates* quantificational recarving.

<sup>9</sup> Originally the paper had a section on what might be considered sentences’ *tertiary* representational task. S\* outperforms S at the tertiary task if its way of carving S’s content does better justice to that content’s *external* nature—its entailment relations with other contents.

## 9. RESPECTING A CONTENT'S IMPLICIT STRUCTURE

What I would give you now, if I had it, is a general analysis of what is involved in a content's being implicitly disjunctive, or negative, or quantificational, and so on through all the logical forms. No such analysis is known to me. On one conception of logical form, I doubt it is even possible. This is the conception whereby a content is disjunctive, say, pure and simple—disjunctive *as opposed to* negative, or existential. There is no reason why some contents shouldn't lend themselves to more than one sort of decomposition.

Because the labels “disjunctive”, “negative,” and so on are apt to sound exclusive (“which is it?”), I will use a slightly altered terminology. Instead of calling a content disjunctive, I will say it has *disjunctivitis*, on the understanding that a content can in principle have two or more –itises at the same time. (The content of “p is an electron or positron” is disjunctive with respect to charge—positive or negative—but conjunctive with respect to charge and mass.)

A different reason for preferring the “–itise” labels is that they suggest not a single defining property, but a cluster of related conditions, not all of which need be present on every occasion. Disjunctivitis (say) might be defined by a largish list of such conditions. Today I will not be trying to finish these lists; it will be enough if we can get them started.

C has *negativitis* iff

it is the complement of a more natural content . . . along with other conditions to be named later.

C has *disjunctivitis* iff

it is the union of a finite number of contents each more natural than it . . . along with other conditions to be named later.

C has *conjunctivitis* iff

it is the complement of a content with disjunctivitis . . . plus other conditions to be named later.

The hard part, of course, is the quantifiers. I will state the proposal, explain it, defend it, and finally apply it.

C has *existentialitis* iff

it is the union of a congruent, complete bunch of contents . . . plus other conditions to be named later.<sup>10</sup>

C has *universalitis* iff

it is the intersection of a congruent, complete bunch of contents . . . plus other conditions to be named later.<sup>11</sup>

<sup>10</sup> One such condition might be that finite sub-unions have disjunctivitis.

<sup>11</sup> One such condition might be that finite sub-intersections have conjunctivitis.

Congruence and completeness are explained as follows. Let the  $C_k$ s be a bunch of contents.

The  $C_k$ s are *congruent* iff the intersection of some of them with one other is more natural than the intersection of some of them with the complement of that other.

The  $C_k$ s are *complete* iff the intersection of some but not all of them is less natural than the intersection of all of them.<sup>12</sup>

A sentence  $P$  mirrors  $C$ 's structure, or, better, mirrors *a* structure of  $C$ , if

- $C$  has negativitis with respect to some content and  $P$  is the negation of a sentence expressing that content, or
- $C$  has disjunctivitis with respect to some contents and  $P$  is a disjunction of sentences expressing those contents, or
- $C$  has conjunctivitis with respect to some contents and  $P$  is a conjunction of sentences expressing those contents, or
- $C$  has existentialitis with respect to some contents and  $P$  is an existential generalization whose instances express those contents, or
- $C$  has universalitis with respect to some contents and  $P$  is a universal generalization whose instances express those contents.

Now the main definition.

$S^*$  does better justice to  $C$  than  $S$  does iff

indent  $S^*$  mirrors a structure of  $C$  that  $S$  fails to mirror.<sup>13</sup>

Note that there is nothing to prevent two sentences' each doing better justice to  $C$  than the other; each mirrors a structure that the other misses.  $S^*$  does *strictly* better justice to  $C$  than  $S$  does if the relation holds in one direction only.

<sup>12</sup> One might wonder how universality can be captured without a "that's all" clause. I have no good answer to this, but here are a couple of thoughts. Sometimes the objects in a domain are "of their own nature" such as to exhaust that domain. The omega-rule testifies to this in the case of numbers; from  $\varphi(1), \varphi(2), \dots, \varphi(k) \dots$ , it directly follows that  $\forall n \varphi(n)$ . That 1, 2, 3, ... are all the numbers is internal to them, obviating the need for a separate "that's all". Another example might be the universal domain (the domain of all possible objects). Perhaps we should think of ourselves as defining universalitis just for this limited case—the case of objects that by nature exhaust their kind. Second, we might amend the definition to say that contents are complete if the intersection of some of them is less natural, not than the intersection of all of them, but than a content "just stronger" than that intersection—in the sense, perhaps, that it properly entails the intersection and all and only its entailments.

<sup>13</sup> "Doing better justice to  $C$ " is understood in this paper as doing better justice to  $C$ 's *internal* nature (= its implicit logical structure). Ideally one would like to assign some weight also to "doing better justice to  $C$ 's external nature" (= its implication relations).

10. “ENUMERATIVE” INDUCTION<sup>14</sup>

Imagine that we are atomic scientists who have never thought of quantifying over numbers. Electrons and protons we have lumped together as “trons.” Whether an atom is electrically charged, we notice, is not always predictable from how many trons it has. Atoms with two trons, or four, or six, are sometimes charged and sometimes neutral. A four-tron atom is neutral if two of its trons are protons and two are electrons; otherwise it is charged. Atoms with one tron, however, or three, or five, are always charged. Further testing reveals that the same holds of atoms with seven trons, or nine. At this point, we stop to review our findings.

[One] Atoms with one tron are charged.

[Two] Atoms with three trons are charged.

...

[Nine] Atoms with seventeen trons are charged.

The data seem to suggest some larger hypothesis. How to express this hypothesis is not clear, until someone has the idea of introducing the device of infinitary conjunction. The data suggest that

$[\infty]$   $\prod_{k = \text{one, two, } \dots}$  Atoms one short of two  $k$  trons are charged.

Not only are [One]–[Nine] evidence for  $[\infty]$ , they confirm it in the way that a lawful generalization is confirmed by its instances, with examined cases supporting unexamined cases. They confirm it in the way a black raven confirms *All ravens are black*, as opposed to the way ~~to~~ a fair coin’s coming up heads confirms *This coin comes up heads every time*.

Statements do not normally provide this kind of support—*inductive* support—to their conjunction. Why now? It is true that this particular conjunction has an especially natural content. But it is unclear why this would make a confirmational difference.

<sup>14</sup> This section was inspired by an example in Field 1998, section 11. Should I believe Charley, when he says there was a foot of snow on the ground in Mobile, Alabama one day in 1936? I should, for he has made many surprising claims in the past, and they all turned out to be true. The inference here appears to be an induction on truth: Charley spoke the truth when he said parts of Virginia are north of parts of New Jersey, Charley spoke the truth when he said the Soviets secretly supplied arms to Chiang Kai-Shek, etc., so most likely Charley is speaking the truth about the snow in Mobile. How are deflationists supposed to understand this inference, rejecting as they do a projectible property of truth? Premises of the form “Charley said ‘ $A_k$ ’ and  $A_k$ ” would not seem to lend inductive support to a conclusion of the form “If Charley said ‘ $B$ ’ then  $B$ .” I adapt Field’s example to the case of numbers. (I should say that Field is less impressed by it than I am, or was when I wrote the paper.)

A familiar if simple-minded picture of confirmation runs as follows. Confirmation is the converse of explanation. Smoke on the mountain indicates the presence of fire there because, and to the extent that, the fire hypothesis explains why there would have been smoke.

- (CON) A given body of data  $D$  confirms hypothesis  $H$  iff
- (a)  $H$  explains  $D$ , or
  - (b)  $H$  follows from an  $H+$  that explains  $D$ .

Say our data  $D$  is various bits of copper all conducting electricity.  $D$  confirms *Copper (as a rule) conducts electricity* because, and to the extent that, *Copper conducts electricity* explains why the tested bits were found to conduct electricity. It confirms *Other, so far unexamined, bits of copper also conduct electricity* because this follows from the explanation we gave of the tested bits' conducting electricity.

To this theory of confirmation, let us now add a simple-minded (broadly Humean) theory of explanation:

- (EXP)  $K$  explains  $D$  iff
- (a)  $K$  entails  $D$
  - (b)  $K$  is a highly natural hypothesis<sup>15</sup>
  - (c)  $K$  is no stronger than (a) and (b) require.<sup>16</sup>

*Copper conducts electricity* is a highly natural data-entailing hypothesis that is weaker than other such hypotheses; it is weaker, e.g., than *Everything conducts electricity*. According to (EXP), then, it is copper's conducting electricity that explains why the observed bits of copper were found to conduct electricity.

Let's now add one further principle—a principle suggested by (CON) and (EXP) and plausible on its face:

- (NAT) if  $D$  confirms  $H$ , then  $D \& H$  is more natural than  $D \& \sim H$ .

The argument from (CON) and (EXP) to (NAT) is as follows. Suppose that  $D$  confirms  $H$ . By (CON),  $H$  either (a) explains  $D$  itself, or (b) follows from an  $H+$  that explains  $D$ . Suppose first that  $H$  explains  $D$  itself. Then  $H$  is the weakest highly natural hypothesis that entails  $D$  (by (EXP)). But then  $D \& \sim H$ , which also entails  $D$ , cannot be highly natural; if it were,  $H$  would not be weakest among highly natural  $D$ -entailing hypotheses.  $D \& H$  is highly natural, though, since  $D \& H = H$  ( $H$  entails  $D$  by (EXP) (a)). So  $D \& \sim H$  is not as natural as  $D \& H$ .

Imagine now that ~~although~~  $H$  does not itself explain  $D$ ;  $H$  follows from  $D$ 's explanation  $H+$ . That  $H+$  explains  $D$  means that  $H+$  is the weakest highly natural hypothesis that entails  $D$  (again by (EXP)). There can be no equally

<sup>15</sup> Meaning, at least, more natural than the data (not necessarily limited to  $D$ ) that  $K$  is called on to explain.

<sup>16</sup> Meaning,  $K$  is weakest among highly natural  $D$ -entailing hypotheses.

natural hypothesis entailing  $D \& \sim H$ , or  $H+$  would not be weakest among highly natural hypotheses entailing  $D$ . Hence  $D \& H$  is implied by a more natural hypothesis (viz.  $H+$ <sup>17</sup>) than any implying  $D \& \sim H$ . Hypotheses are more or less natural, one assumes, according to the naturalness of what implies them; so  $D \& H$ , being implied by  $H+$ , is more natural than  $D \& \sim H$ . This completes the argument that  $D$  combines more naturally with an  $H$  that it confirms than with that  $H$ 's negation.

Now, as we saw,  $[\infty]$  bears the same inductive relations to  $[\text{One}]$ ,  $[\text{Two}]$ , . . . , and  $[\text{Nine}]$  as a lawlike generalization bears to its observed instances. How, according to our simple-minded theory of confirmation, must  $[\text{One}]$ – $[\text{Nine}]$  ( $= D$ ) and  $[\infty]$  ( $= H$ ) be related for this to be so? A lawlike generalization's observed instances are supposed to confirm

- (i) the generalization as a whole,

and

- (ii) the generalization's unobserved instances.<sup>18</sup>

The question, then, is this: How must  $D$  and  $H$  be related for  $D$  to confirm

- (i')  $H$  as a whole, that is,  $\prod_{k = \text{one, two, } \dots} [k]$ ,

and also

- (ii')  $H$ 's unobserved instances, that is, *Atoms with nineteen trons are charged*, *Atoms with twenty-one trons are charged*, etc.

It follows from (NAT) that  $D$  confirms  $H$ 's unobserved instances only if

- (ii'')  $D$ 's conjunction with  $[\text{Ten}]$  ( $[\text{Eleven}]$ ,  $[\text{Twelve}]$ , . . . ) is more natural than its conjunction with not- $[\text{Ten}]$  (not- $[\text{Eleven}]$ , not- $[\text{Twelve}]$ , etc)

It follows from (CON) and (EXP) that  $D$  confirms  $H$  as a whole only if

- (i'')  $H$  is weakest among highly natural  $D$ -entailing hypotheses.

This is interesting because (ii'') is a special case of what we above called Congruence.<sup>19</sup>  $H$ 's conjuncts are congruent provided that (ii'') continues to hold no matter which  $[k]$ 's go in for  $[\text{One}]$ – $[\text{Nine}]$  in  $D$ . Similarly (i''), assuming it too continues to hold for other choices of  $D$ , implies what we above called Completeness. For suppose the conjunction of only *some* of  $[\text{One}]$ ,  $[\text{Two}]$ ,  $[\text{Three}]$ , . . . ,  $[\text{Nine}]$ ,  $[\text{Ten}]$ ,  $[\text{Eleven}]$ , . . . was as natural as the conjunction of

<sup>17</sup> Recall that  $H+$  implies  $H$ .

<sup>18</sup> A fair coin's coming up heads confirms (makes it likelier) that the coin always comes up heads. But it doesn't *inductively* confirm that the coin always comes up heads, because it doesn't make it any likelier that it will come up heads on the next toss.

<sup>19</sup> I am fudging here the distinction between more or less natural *hypotheses* and hypotheses with more or less natural contents.



all of them. Then, letting that sub-conjunction be D, H would not be weakest among highly natural D-entailing hypotheses; D would itself be a weaker such hypothesis. H's conjuncts are complete provided that (i') continues to hold no matter which [k]'s go in for [One]–[Nine] in D.

Where does this leave us? H's conjuncts inductively confirm H only if they are complete and congruent. For them to be complete and congruent is for H's content to have universalitis with respect to its conjuncts' contents. So,

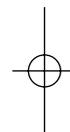
Statements like *Atoms with three trons are charged* inductively confirm *Atoms with one or three or five or . . . trons are charged* only if the latter's content has universalitis with respect to the contents of statements like the former.<sup>20</sup>

Contents with universalitis are contents that invite recarving as universal generalizations. To think of [ $\infty$ ] as inductively supported by its conjuncts, then, is to conceive [ $\infty$ ]’s content as deeply, underlyingly, *general*, whatever grammatical face it has been presenting to the world up to now.

By now the neo-platonic drift of all this should be clear. A generalization needs a domain to generalize over. That domain would seem, in this case, to be the natural numbers. The content-respecting way to express [ $\infty$ ] is

$\forall a \forall n$  (if the number of  $a$ 's trons is  $2n - 1$ , then  $a$  is charged).

This gives new meaning to the term “enumerative induction.” Numbers worm their way into empirical contents not just on representational grounds, but also evidential ones. Induction needs an axis to operate along. Numbers provide that axis.<sup>21</sup>



## 11. MORALS IF ANY

The hypothesis of this paper is that S invites recarving by S\* iff S\* does better justice than S to C's implicit logical structure. The question should finally be raised of how far this hypothesis supports the program of Fregean platonism. I confess I don't know. One complication is that the Fregean platonist may well look askance at our ontology of cases and contents—not because abstract ontology is objectionable per se, but because it is supposed to be a consequence of the project rather than a presupposition of it.

<sup>20</sup> H's content is, as ever, the conjunction of the contents of [One], [Two], . . . It thus *admits* conjunctive recarving. The present issue is what sort of recarving H's content *invites*. ⊙

<sup>21</sup> Is this to say that numbers are evidentially relevant—that they bear evidentially on empirical events? A similar question is sometimes raised about causal relevance. How did the missile get through our air defense system? It was surrounded by a large number of decoys. The large number is not as such causally active; it merely marks the fact that there were lots of decoys. Numbers' role in induction is deeper than this. It informs the relation itself, not just one of its relata. But the general point is the same. Facilitating a relation is not the same as standing in that relation.



But let's suppose that that can somehow be dealt with. Does it help Fregean platonism if the content of "there are as many Fs as Gs" invites recarving as "the number of Fs = the number of Gs"?<sup>22</sup>

On the one hand, I want to say Yes. If a carving-based reason is to be given why we should (or can at no epistemological risk, or already implicitly do) countenance numbers,<sup>23</sup>

( $\diamond$ ) known truths—contents we know to be true—can be rearticulated in a numerical fashion

seems a good deal less convincing than

( $\square$ ) known truths—contents we know to be true—demand to be rearticulated in numerical fashion.

As we saw, ( $\diamond$ ) makes an exceedingly weak case for numbers, no better than could be made for duds, or whereabouts, or party affiliations. ( $\square$ ) makes a much stronger case. Known truths do not cry out to be rearticulated in terms of duds.

The case ( $\diamond$ ) makes for numbers is also un-Fregean. Numbers are supposed to be discernible in *cardinality* contents specifically—the kind found, for instance, on the right hand side of Hume's Principle. But all kinds of contents can be rearticulated so as to involve numbers. Cardinality contents have no particular advantage.

( $\square$ ) avoids this problem. Numbers are not invited into the content of any old truth (e.g., that Hume reckons on a speedy dissolution), but only truths whose implicit logical structure is thereby illuminated.

( $\square$ ) makes the better case. But the case is not irresistible. That known truths cry out for numerical rearticulation could be heard less as a theoretical argument for the objects' existence, than a practical argument for postulating them quite regardless of whether they exist. (If numbers did not exist, it would be necessary to invent them.) It can even be heard as a reason to suspect that they *already have been* postulated regardless of whether they are there. This way lies fictionalism, or figuralism, or presuppositionalism, presumably of the hermeneutic variety.

I take no stand here on which of these views—Fregean platonism or hermeneutic fictionalism/figuralism/presuppositionalism—is in the end preferable. My point is directed at both equally. Both camps try to portray number-talk as harmless and unobjectionable. Our entitlement to arithmetic is secure because it flies beneath skeptical radar; there is nothing in it to arouse intellectual opposition. I am suggesting that both camps would do better to take the offensive, arguing that the facts as we know them cry out for numerical treatment and we ought to

<sup>22</sup> To be sure, I haven't directly argued that it does. A key lemma would be that the "the number of Fs = the number of Gs" does better justice than "there are as many Fs as Gs" to their shared content's external nature (implication relations).

<sup>23</sup> I say "*if* a carving-based reason is to be given" because Crispin Wright offers a different sort of reason; the key point for him is that Hume's Principle canonically explains the concept of number.

heed their call. It can be left as a further question whether the heeding should take the form of believing in numbers, or acting as if we believed in them.

## REFERENCES

- Burge, T. (1998). "Frege on Knowing the Foundation." *Mind* 107: 305–347.
- (1992). "Frege on Knowing the Third Realm." *Mind* 101: 633–650.
- (1984). "Frege on Extensions of Concepts, from 1884 to 1903." *Philosophical Review* 93: 3–34.
- Davidson, D. (1979). "The Inscrutability of Reference." *Southwestern Journal of Philosophy* 10: 7–19.
- Eklund, M. (2006). "Neo-Fregean Ontology." *Philosophical Perspectives*, 20 (1): 95–121.
- Field, H. (1994). "Deflationist Views of Meaning and Content." *Mind* 103: 249–285.
- Frege, G. (1997). *The Frege Reader*. Oxford and Malden, MA: Blackwell Publishers.
- Hale, B. (2001). "A Response to Potter and Smiley: Abstraction by Recarving." *Proceedings of the Aristotelian Society* 101: 339–358.
- (1997). "Grundlagen § 64." *Proceedings of the Aristotelian Society* 97: 243–61.
- (1988). *Abstract Objects*. New York: Blackwell.
- Hale, B. and C. Wright (2001). *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Clarendon Press.
- Heck Jr, R. G. (2000). "Syntactic Reductionism." *Philosophia Mathematica* 8 (2): 124–149.
- Levine, J. (1996). "Logic and Truth in Frege: II." *Aristotelian Society Supp.* Vol. 70: 141–175.
- Lewis, D. and R. Langton (2001). "Redefining 'Intrinsic'." *Philosophy and Phenomenological Research* 63 (2): 381–398.
- Potter, M. and T. Smiley (2001). "Abstraction by Recarving." *Proceedings of the Aristotelian Society* 101: 327–338.
- (2002). "Recarving Content: Hale's Final Proposal." *Proceedings of the Aristotelian Society* 102: 351–354.
- Rosen, G. and J. P. Burgess (2005). "Nominalism Reconsidered," in *The Oxford Handbook of Philosophy of Mathematics and Logic*, ed. Stewart Shapiro. Oxford: Oxford University Press: 515–535.
- Wright, C. (1983). *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.
- Yablo, S. (2006). "Non-Catastrophic Presupposition Failure," in *Content and Modality: Themes from the Philosophy of Robert Stalnaker*, ed. Judith Thomson and Alex Byrne. Oxford: Oxford University Press. [Chapter 11 in this volume]
- (2005). "The Myth of the Seven," in Kalderon, ed., *Fictionalist Approaches to Metaphysics*, Oxford: Oxford University Press [Chapter 9 in this volume].
- (2002). "Abstract Objects: A Case Study." *Noûs Supplement* 12: 220–240 [Chapter 8 in this volume].
- (2001). "Go Figure: A Path through Fictionalism." *Midwest Studies in Philosophy* 25: 72–102 [Chapter 7 in this volume].

## Non-Catastrophic Presupposition Failure

### 1. BACKGROUND

I will be talking in this paper about the problem of presupposition failure. The claim will be (exaggerating some for effect) that there is no such problem—more like an *opportunity* of which natural language takes extensive advantage.

The last two sentences are a case in point. The first was, “I am going to talk about the F”; the second was, “there is no F.” If the second sentence is true—there is no F—then the first sentence, which presupposes that there is an F, suffers from presupposition failure. In theory, then, it should strike us as somehow compromised or undermined. Yet it doesn’t. So here is one case at least where presupposition failure is not a problem.

The title is meant to be understood compositionally. *Presuppositions* are propositions assumed to be true when a sentence is uttered, against the background of which the sentence is to be understood. Presupposition *failure* occurs when the proposition assumed to be true is in fact false.<sup>1</sup> Failure is *catastrophic* if it prevents a thing from performing its primary task, in this case making an (evaluable) claim. Non-catastrophic presupposition failure then becomes the phenomenon of a sentence still making an evaluable claim despite presupposing a falsehood.

I said that presuppositions were propositions taken for granted when a sentence is uttered, against the background of which the sentence is to be understood.<sup>2</sup>

Papers sort of like this one were presented at Indiana, UC Davis, UC Berkeley, UC San Diego, Yale, Brown, Penn, Kentucky, Oxford (as the 2005 Gareth Evans lecture), ANU, Monash, the Chapel Hill Colloquium, an APA session on Metaontology, and graduate student conferences at Pittsburgh and Boulder. I am grateful to Richard Holton, Sally Haslanger, Agustin Rayo, Caspar Hare, Kai von Fintel, Danny Fox, Irene Heim, Karen Bennett (who commented at the APA), Anne Bezuidenhout (who commented at Chapel Hill), Larry Horn, Sarah Moss, John Hawthorne, Lloyd Humberstone, and especially Bob Stalnaker for questions and advice. I learned too late of the literature on “logical subtraction” (see Humberstone 2000 and references there); it holds out hope of a different and perhaps more straightforward route to incremental content.

<sup>1</sup> Really I should say “untrue” rather than “false,” to allow for presuppositions that lack truth-value because they themselves suffer from presupposition failure. Looking ahead to the Donnellan examples, “The man drinking a martini is *that* guy” is (so it seems) not false but undefined if no one is drinking a martini.

<sup>2</sup> Are we to think of presupposition as a relation that *sentences* bear to propositions (Strawson), or a relation that *speakers* bear to propositions (Stalnaker)? There may be less of a difference here than meets the eye. The first relatum for Strawson is *utterances* or tokens of *S*, from which it is a

It would be good to have some tests for this. Here are three, loosely adapted from the paper that got me thinking about these issues (von Fintel 2004—don't miss it!).<sup>3</sup>

One is the “Hey, wait a minute” test.<sup>4</sup> If  $\pi$  is presupposed by  $S$ , then it makes sense for an audience previously unaware of  $\pi$  to respond to an utterance of  $S$  by saying “Hey, wait a minute, I didn't know that  $\pi$ .” If  $\pi$  were asserted, that response would be silly; of course you didn't know, the point of uttering  $S$  was to tell you. Suppose you say, “I'm picking my guru up at the airport.” I can reply, “Hey, I didn't know you had a guru,” but not, “Hey, I didn't know you were going to the airport.” This suggests that your having a guru was presupposed while your going to the airport was asserted. A likelier response to what is asserted is, “Is that so, thanks for telling me.”<sup>5,6</sup>

Second is the attitude attributed when we say that someone denies that  $S$ , or hopes or regrets that  $S$ ; the presupposition  $\pi$  is exempted from the content of that attitude. Hoping you will pick up your guru at the airport may be in part hoping your guru will be picked up, but it is not hoping that you have a guru in the first place. Denying that you are going to pick up your guru at the airport is not denying the conjunction of *you have a guru* with *you are going to pick your guru up at the airport*.<sup>7</sup> So a second mark of presuppositions is that  $\pi$  does not

short step to speakers presupposing this or that *in uttering*  $S$ . Stalnaker for his part appreciates that certain sentences  $S$  should not be uttered unless this or that is (or will be as a result of the utterance) pragmatically presupposed. It does little violence to either's position to treat “ $S$  presupposes  $\pi$ ” as short for “All (or most, or contextually salient) utterances of  $S$  presuppose  $\pi$ ,” and that in turn as short for “Speakers in making those utterances always (often, etc.) presuppose that  $\pi$ .” (Von Fintel ms and Simons 2003 are illuminating discussions.) Semantic presupposition would be the special case of this where  $S$  presupposes  $\pi$  as a matter of meaning, that is,  $S$ -users presuppose  $\pi$  not for conversational reasons but because semantic rules require it.

<sup>3</sup> Strawson noticed that while some King-of-France sentences strike us as unevaluable (“The KoF is bald”), others seem false (“The KoF visited the Exhibition yesterday”). Von Fintel criticizes earlier accounts of this contrast (by Strawson and Peter Laserson) and proposes an interesting new account. He does not address himself to a third possibility noted by Strawson, that a sentence with false presuppositions should strike us as true. This paper agrees with von Fintel's basic idea: some KoF-sentences “are rejected as false . . . because they misdescribe the world in two ways: their presupposition is false, but in addition there is another untruth, which is *independent* of the failure of their presupposition” (2004, 325). But it implements the idea differently.

<sup>4</sup> Taken apparently from Shanon 1976.

<sup>5</sup> This test seems to work best for semantic presuppositions (see note 3). Looking ahead a bit, “The man drinking a martini is a philosopher” does not invite the reply, “Hey, I didn't know *that* guy was the one drinking a martini.” One can, however, say, “Hey, I didn't know that guy was drinking a martini.” So perhaps a version or variant of the test applies to (some) non-semantic presuppositions as well.

<sup>6</sup> Von Fintel attributes to Percus a test that is in some ways similar. “ $R$ , and what's more,  $S$ ” sounds fine if  $S$  asserts more than  $R$ , but wrong if  $S$  only presupposes more. So, “John thinks Judy is a chiropractor” can be followed by “And what's more, he is right to think Judy is a chiropractor,” but not “And what's more, he realizes Judy is a chiropractor.” This seems to indicate that “He realizes that BLAH” presupposes what “He is right to think that BLAH” asserts, viz. that BLAH, and asserts what it presupposes, viz. that he believes that BLAH.

<sup>7</sup> This observation goes essentially back to Frege. Frege considers the sentence “Whoever discovered the elliptic form of the planetary orbits died in misery.” He notes that its negation is

figure in what you hope or deny or regret in hoping or denying or regretting that  $S$  (Stalnaker 1999, 39).

A third test is that presuppositions within limits *project*, that is,  $\pi$  continues to be presupposed by more complex sentences with  $S$  as a part. If you say, “I don’t have to pick up my guru after all,” or, “It could be I will have to pick my guru up,” these statements still intuitively take it for granted that you have a guru. Our earlier tests confirm this intuition. One can still reply, “Hold on a minute, you have a guru?” And to hope that you don’t have to pick your guru up is not to hope that you have a guru.<sup>8</sup>

Note that one test sometimes used to identify presuppositions is missing from this list:  $\pi$  is presupposed iff unless  $\pi$  holds,  $S$  says *nothing true or false*. That test is useless in the present context because it makes NCPF impossible;  $\pi$  is not classified as a presupposition unless its failure would be catastrophic.

A sentence suffers from catastrophic presupposition failure only if, as Strawson puts it, “the whole assertive enterprise is wrecked by the failure of [ $S$ ’s] presupposition” (1964, 84). There is also the phenomenon of what might be called *disruptive* presupposition failure. This occurs when  $\pi$ ’s failure does not wreck the assertive enterprise so much as reveal it to have been ill advised. It could be, for instance, that  $\pi$  was an important part of the speaker’s *evidence* for  $S$ . It could be that  $\pi$  was part of what made  $S$  *relevant* to the rest of the conversation. It could even be that  $S$  *entails*  $\pi$  so that  $\pi$ ’s falsity guarantees that  $S$  is false too.<sup>9</sup>

Disruption is bad, but it is not (in our sense) a catastrophe. On the contrary, a remark is implausible or irrelevant or false because of what it says, and that something was said suggests that the assertive enterprise has not been wrecked after all. I mention this because Stalnaker, who has written the most about these topics, is addressing himself more often to the disruptive/non-disruptive distinction than the catastrophic/non-catastrophic distinction.<sup>10</sup> This paper is meant to be entirely about the latter.

“Whoever etc. did not die in misery” rather than “Either whoever discovered the elliptic form of the planetary orbits did not die in misery or there was nobody who discovered the elliptic form of the planetary orbits” (1872, 162–3). If we assume (as he did) that denial is assertion of the negation, this amounts to the claim that “Somebody discovered the elliptic form of the planetary orbits” is no part of what is denied when we deny that “whoever etc. died in misery.”

<sup>8</sup> Presuppositions are further distinguished by the way they fail to project in certain contexts, such as conditionals with  $\pi$  as antecedent. “I don’t remember if I have a guru, but if I do, I should remember to ask for a new mantra” does not presuppose that I have a guru.

<sup>9</sup> This relates to a passage in “Pragmatic Presuppositions”: “Using the pragmatic account [of presupposition], one may say that sometimes when a presupposition is required by the making of a statement, what is presupposed is also entailed, and sometimes it is not. One can say that ‘Sam realizes that  $P$ ’ entails that  $P$ —the claim is false unless  $P$  is true. ‘Sam does not realize that  $P$ ,’ however, does not entail that  $P$ . That proposition may be true even when  $P$  is false. All this is compatible with the claim that one is required to presuppose that  $P$  whenever one asserts or denies that Sam realizes it” (1999, 54).

<sup>10</sup> For instance here: “Where [presuppositions] turn out to be false, sometimes the whole point of the inquiry, deliberation, lecture, debate, command, or promise is destroyed, but at other times it

≡ caps

## 2. RELEVANCE TO PHILOSOPHY

Why should we care about non-catastrophic presupposition failure? There are reasons from the philosophy of language, from epistemology, and from metaphysics.

The philosophy of language reason is simple. All of the best-known theories of presupposition (among philosophers, anyway) suggest that failures are or ought to be catastrophic. This is clearest for Frege's and Strawson's theories—<sup>↗</sup>for those theories more or less *define* a sentence's presuppositions as preconditions of its making an evaluable claim. Assuming as before that a sentence's primary task is to offer a true or false account of how things are, presuppositions on Frege's and Strawson's theories are *automatically* propositions whose failure has catastrophic effects.

Next consider Stalnaker's theory of presupposition. Stalnaker-presupposition is in the first instance a relation between speakers and propositions; one presupposes  $\pi$  in uttering  $S$  if one thinks that  $\pi$  is (or will be, as a result of the utterance) common ground between relevant parties. A *sentence* presupposes  $\pi$  only to the extent that  $S$  is not appropriately uttered unless the speaker presupposes that  $\pi$ .

Why on this account should presupposition failure be problematic? Well, the point of uttering  $S$  is to draw a line through the set of worlds still in play at a particular point in the conversation—one is saying that *our* world is on the  $S$ -true side of the line rather than the side where  $S$  is false. Since the worlds still in play are the ones satisfying all operative presuppositions, the speaker by presupposing  $\pi$  is arranging things so that her remark draws a line through the  $\pi$ -worlds only.

But then what happens when  $\pi$  is false? Because the actual world is outside the region through which the line is drawn, it is hard to see how in drawing this line the speaker is saying anything about actuality. It's as though I tried to locate Sicily for you by saying that *as between North and South Dakota*, it's in the South, although truth be told it's not in either Dakota. Similarly it is not clear how I can locate actuality for you by saying that as between the  $\pi$ -worlds where  $S$  is true and the ones where it is false, it's in the first group, although truth be told it's not a  $\pi$ -world at all.<sup>11</sup>

That was the philosophy of language reason for caring about NCPF; the standard theories seem to rule it out. A much briefer word now on the epistemological and metaphysical reasons.

does not matter much at all . . . Suppose . . . we are discussing whether we ought to vote for Daniels or O'Leary for President, presupposing that they are the Democratic and Republican candidates respectively. If our real interest is in coming to a decision about who to vote for . . . , then the debate will seem a waste of time when we discover that in reality, the candidates are Nixon and Muskie. However if our real concern is with the relative merits of the character and executive ability of Daniels and O'Leary, then our false presupposition makes little difference" (1999, 39).

<sup>11</sup> See Beaver 2001 for theories of the kind favored by many linguists. These seem at least as unaccommodating of NCPF as the ones philosophers like, for a reason noted by Simons: "Dynamic theories of presupposition claim that presupposition failure results in undefinedness of the context update function—the dynamic correlate of truth valuelessness" (2003, 273).

The epistemological reason has to do with testimony, or learning from others. Someone who utters a sentence  $S$  with truth-conditions  $C$  ( $S$  is true if and only if  $C$  obtains) might seem to be telling us that  $C$  *does* obtain. But if we bear in mind that  $\pi$  is one of the conditions of  $S$ 's truth, we see that that cannot be right. For it makes two false predictions about the phenomenon of NCPF. The first is that *all presupposition failure is non-catastrophic*; if  $\pi$  is false, then the speaker is *telling* us something false, hence the assertive enterprise has not been wrecked. The second is that what the speaker is telling us *can never be true*. The fact is that some presupposition failure is catastrophic and some isn't; and the claim made can be either true or false. To suppose that speakers are saying inter alia that  $\pi$  in uttering  $S$  collapses the first two categories—catastrophic, non-catastrophically true—into the last—non-catastrophically false.

So here is the epistemological relevance of NCPF. It reminds us that speakers are not in general vouching for *everything* the truth of their sentence requires; they vouch for the asserted part but not (in general) for the presupposed part.

This leads to the metaphysical reason for caring about NCPF. Quine famously argues like so: "Scientists tell us that the number of planets is 9; that can't be true unless there are numbers; so scientists tell us inter alia that there are numbers; so unless we consider ourselves smarter than scientists, we should believe in numbers." This assumes that speakers are vouching for *all* the truth-conditions of the sentences coming out of their mouths. But there being a thing that numbers the planets is no part of what Clyde Tombaugh (the discoverer of Pluto) was telling us—no part of what he was giving his professional opinion about—when he spoke the words, "The number of planets is 9." A different metaphysical upshot will be mentioned briefly at the end.

### 3. FREGE AND STRAWSON

I said that the best-known theories suggest that all presupposition failure ought to be catastrophic, and that the suggestion is implausible. I did not say that the best-known theorists are unaware of this problem. Well, Frege might have been unaware of it. Even he, though, gives an example that might be taken as a case in point: "Somebody using the sentence 'Alfred has still not come' actually says 'Alfred has not come,' and at the same time hints—but only hints—that Alfred's arrival is expected. Nobody can say: 'since Alfred's arrival is not expected, the sense of the sentence is false' " (1918, 331).

Frege's use of *hint* makes it sound as though we are dealing with an implicature. But "still" is by the usual tests a presupposition trigger. ("Hang on, I didn't know Alfred was supposed to be here!") Suppose for argument's sake that the tests are right.

Frege says that the thought is not automatically false if Alfred was unexpected. By this he presumably means that the thought's truth-value depends not on

how expected Alfred was but on whether he has indeed come. Even if the presupposition fails—he was *not* expected—a claim is still made that can be evaluated as true or false.

So the Alfred example looks like a case of non-catastrophic presupposition failure. Of course, Frege would not see it that way, because the presuppositions that he (and later Strawson) has mainly in mind are *existential* presuppositions: “If anything is asserted there is always an obvious presupposition that the simple or compound proper names used have reference” (1872, 162).

The sentence “Whoever discovered the elliptic form of the planetary orbits died in misery” is said to lack truth-value unless someone did indeed make the indicated discovery (1872, 162). Strawson in similar fashion says that if someone produced the words “The King of France is bald,” we would be apt to say that “the question of whether his statement was true or false *simply did not arise*, because there was no such person as the King of France” (1950, 12).

But, and here he goes beyond Frege, Strawson notices that failure even of a sentence’s existential presuppositions does not prevent it from making an evaluable claim:

Suppose, for example, that I am trying to sell something and say to a prospective purchaser *The lodger next door has offered me twice that sum*, when there is no lodger next door and I know this. It would seem perfectly correct for the prospective purchaser to reply *That’s false*, and to give as his reason that there was no lodger next door. And it would indeed be a lame defense for me to say, *Well, it’s not actually false, because, you see, since there’s no such person, the question of truth and falsity doesn’t arise*. (1954, 225)./

This is an example of what Strawson calls “radical failure of the existence presupposition” (1964, 81), radical in that “there just is no such particular item at all” as the speaker purports to be talking about. It shows that for the existence presupposition to fail radically is not necessarily for it to fail catastrophically.

Now, if the existence presupposition can fail radically—there is no such item as the speaker purports to be talking about—one expects that the uniqueness presupposition could fail radically too—there are *several* items of the type the speaker purports to be talking about. Consider another example of Strawson’s:

if, in Oxford, I declared, “The Waynflete Professor of Logic is older than I am,” it would be natural to describe the situation by saying that I had confused the titles of two Oxford professors [Waynflete Professor of Metaphysics and Wykeham Professor of Logic], but whichever one I meant, what I said about him was true. (1954, 227)./

This becomes *radical* failure of the uniqueness presupposition if we suppose that in confusing the titles, Strawson had confused the individuals too, so that his remark was no more directed at the one than the other. Does the failure thus reconstrued remain *non-catastrophic*? I think it does. The remark strikes us as

false if the Waynflete and Wykeham Professors are both younger than Strawson, and true (or anyway truer) if he is younger than them.<sup>12</sup>

What about *non-radical* failure of the existential and uniqueness presuppositions? By a *non-radical* failure I mean that although the description used is not uniquely satisfied, the subject *does* have a particular item in mind as the intended referent. The uniqueness presupposition fails non-radically when one says, “The square root of N is irrational,” meaning to refer to the positive square root, forgetting or ignoring that N has a negative root too. This kind of remark does not court catastrophe since it strikes us as correct if both roots are irrational, and incorrect if both are rational, and no other outcome is possible.

That was my example of non-radical failure of the uniqueness presupposition, not Strawson’s; his would be the Oxford mix-up, assuming that the intended referent was, say, Gilbert Ryle, then Waynflete Professor of Metaphysics. Strawson also gives an example where it is the existential presupposition that non-radically fails:

perhaps, if I say, “The United States Chamber of Deputies contains representatives of two major parties,” I shall be allowed to have said something true even if I have used the wrong title, a title, in fact, which applies to nothing. (1954, 227)<sup>13</sup>

So although Strawson doesn’t put it this way, his discussion suggests a four-fold classification along the lines shown in Table 12.1.<sup>14</sup> The fourth of Strawson’s

Table 11.1

	<i>uniqueness</i> presupposition	<i>existential</i> presupposition
radical failure of the	Waynflete Prof of Logic <sup>15</sup>	lodger next door
non-radical failure of the	square root of N	Chamber of Deputies

<sup>12</sup> Suppose Strawson had said, “The Philosophy Professor at St Andrews is older than me,” not realizing that St Andrews had two professors. Such a statement again seems correct if both are older and incorrect if both are younger—indeed (arguably) if either is younger. Stalnaker in conversation suggests treating this as a case of pragmatic ambiguity; the utterance seems true when it is true on both disambiguations, false when it is false on both (or perhaps false on either). I do not see how to extend this treatment to superficially similar cases. “All eight solar planets are inhabited” seems false, but it is presumably not ambiguous between nine attributions of inhabitedness, each to all solar planets but one.

<sup>13</sup> This example is important in Strawson’s debate with Russell. Some empty-description sentences strike us as false, as Russell’s semantics predicts. But others are such that “if forced to choose between calling what was said true or false, we shall be more inclined to call it true” (Strawson 1954, 227). Russell cannot claim too much credit for plugging truth-value gaps, if he sometimes plugs in the wrong value. (I ignore the wide-scope negation strategy as irrelevant to the examples Strawson is concerned with here.)

<sup>14</sup> This classification is not meant to be exhaustive; perhaps, e.g., the description applies to exactly one thing, but that thing is not the intended referent.

<sup>15</sup> Understood so that the speaker is thinking confusedly of both professors at once.

categories—non-radical failure of the existential pre-supposition—proved the most influential, as we shall see.

#### 4. DONNELLAN AND STALNAKER

Strawson appreciates, of course, that the judgments just noted seem at odds with his official theory, particularly with the principle that “If someone asserts that the  $\phi$  is  $\psi$  he has not made a true or false statement if there is no  $\phi$ ” (Donnellan 1966, 294). Donnellan’s famous counterexample to that principle would thus not have come as a surprise to him:

Suppose one is at a cocktail party and, seeing an interesting-looking person holding a martini glass, one asks, “Who is the man drinking a martini?” If it should turn out that there is only water in the glass, one has nevertheless asked a question about a particular person, a question it is possible for someone to answer (1966, 287).

Given that “Strawson admits that we do not always refuse to ascribe truth to what a person says when the definite description he uses fails to fit anything (or fits more than one thing)” (1966, 294), what does Donnellan think he is adding to Strawson’s own self-criticism? Donnellan is not very explicit about this, but here is my best guess as to his reply.

What Strawson admits is that the person has said *something* true. He does not (according to Donnellan) admit that the statement *originally at issue*, viz. “the man drinking a martini is a famous philosopher” is true. One might wonder, of course, what we are doing if not “awarding a truth value. . . to the original statement.” The answer is that we “amend the statement in accordance with [the speaker’s] guessed intentions and assess the amended statement for truth or falsity” (Strawson 1954, 230). The statement Strawson is willing to call true, then, is not the one suffering from presupposition failure, and the one suffering from presupposition failure he is not willing to call true. (Elsewhere Strawson says the original statement is true only in a *secondary* sense.) Donnellan is bolder: he thinks that the *unamended*, original statement “The  $\phi$  is  $\psi$ ” can be true in the absence of  $\phi$ s, if the description is used referentially.

A second difference between Donnellan and Strawson is this. Strawson paints a mixed picture featuring on the one hand a *presupposition* that the description is uniquely satisfied, and on the other hand an *intention to refer* with that description to a certain object. Donnellan simplifies matters by turning the referential intention into an additional presupposition:

[W]hen a definite description is used referentially, not only is there in some sense a presupposition . . . that someone or something fits the description, . . . but there is also a quite different presupposition; the speaker presupposes of some *particular* someone or something that he or it fits the description. In asking, for example, “Who is the man drinking a martini?” where we mean to ask a question about that man over there, we are

presupposing that that man over there is drinking a martini—not just that *someone* is a man drinking a martini (1966, 289).

This may not seem like progress; before we had one failed presupposition to deal with, now we have two. But, and this is the third difference between Donnellan and Strawson, the “new” failed presupposition, rather than being an obstacle to evaluation, is what *enables* evaluation, by pointing the way to an evaluable hypothesis: that man is a famous philosopher.

Stalnaker attempts to put all this on a firmer theoretical foundation. Imagine O’Leary saying, “The man in the purple turtleneck is bald,” where it is understood that the man in question is *that* man (Daniels). The propositional content of O’Leary’s statement is that Daniels is bald. The fixation of content here is along lines more or less familiar from Kaplan. Just as the character of an expression like “you” determines its denotation as a function of context, “there are relatively systematic rules for matching up [referential] definite descriptions with their denotations in a context” (1999, 41). The rule for “you” is that it contributes the addressee; the rule for a referential description is that it contributes “the one and only one member of the appropriate domain who is presupposed to have the property expressed in the description” (1999, 41). Crucially from our perspective,

it makes no difference whether that presupposition is true or false. The presupposition helps to determine the proposition expressed, but once that proposition is determined, it can stand alone. The fact that Daniels is bald in no way depends on the color of his shirt (1999, 43).

So we see that Stalnaker does have an account to offer of *some* cases of NCPF. NCPF occurs (in these cases) for basically Kaplanian reasons. A conventional meaning is given by a systematic character function mapping contexts (= sets of worlds) to propositions. And there is nothing to stop a set of worlds from being mapped to a proposition defined on worlds outside of the set.

This is fine as far as it goes. But NCPF is ubiquitous, and character as Kaplan understands it is reserved to a few special terms. Stalnaker knows this better than anyone, of course; he was one of the first to charge two-dimensionalists with an undue optimism about the project of extending Kaplan-style semantics from demonstratives to the larger language. Some NCPF may be a matter of characters mapping contexts to propositions defined outside those contexts, but not much. An example of Kripke’s brings out the extent of the difficulty:

Two people see Smith in the distance and mistake him for Jones. They have a brief colloquy: “What is Jones doing?” “Raking the leaves.” “Jones,” in the common language of both, is a name of Jones; it *never* names Smith. Yet, in some sense, on this occasion, clearly both participants in this dialogue have referred to Smith, and the second participant has said something true about the man he referred to if and only if Smith was raking the leaves. (Kripke 1977, 14)

Assuming Smith was raking the leaves, the second participant says something true with the words, “Jones is raking the leaves,” despite (or because of) the false presupposition that it is Jones they see off in the distance. The example has a Donnellan-like flavor, but the explanation will have to be different; a proper name like “Jones” does not have a reading on which it denotes whoever is presupposed to be Jones in the relevant context. This is why I say there is no general account of NCPF in Stalnaker.<sup>16</sup> I will be suggesting, however, that he does provide the materials for such an account.

So, to review. A sentence’s presuppositions are (generally) no part of what it says. Presuppositions can however function as *determinants* of what is said. The suggestion is that they can influence what is said equally well even if false. It remains to explain how exactly the trick is pulled off. Explaining this will be difficult without an account of the mechanism by which presuppositions exert their influence. Because we are really asking about that mechanism. Does it ever, in the course of its  $\pi$ -induced operations, find itself wondering whether  $\pi$  is true?

There are hints in the literature of three strategies for making  $\pi$  (not a part of but) a guide to asserted content. The first tries to get at what  $S$  says by *ignoring* the possibility that  $\pi$  fails. The second tries to get at what  $S$  says by *restoring*  $\pi$  when it does fail. The third tries to get at what  $S$  says by asking what *more* than  $\pi$  needs to be true for  $S$  to be true. I will be arguing against IGNORE and RESTORE and defending SAY-MORE.

## 5. IGNORE

Asserted content as conceived by the first strategy addresses itself only to  $\pi$ -worlds. It just ignores worlds where  $\pi$  fails. Thinking of contents as functions from worlds to truth-values, ignoring a world is being undefined on that world.  $S$ ’s asserted content is thus a partial function mapping  $\pi$ & $S$ -worlds to truth,  $\pi$ & $\sim S$ -worlds to falsity, and worlds where  $\pi$  fails to nothing at all.

- [1]  $S$ ’s asserted content  $S$  is the proposition that is true (false) in a  $\pi$ -world  $w$  iff  $S$  is true (false) in  $w$ , and is otherwise undefined.<sup>17</sup>

There might seem to be support for this in a passage from Stalnaker:

<sup>16</sup> This is not to say he doesn’t have explanations to offer in particular cases. Often he appeals to a device like Strawson’s (see above). The original statement—“Jones is raking the leaves”—suffers from presupposition failure, so is not evaluable. Had the speaker been better informed, she would have made a statement—“Smith is raking the leaves”—whose presuppositions are true. Our evaluation of the second statement is then projected back onto the first.

<sup>17</sup> ~~So far~~ this says nothing about  $S$ ’s truth-value in worlds where  $\pi$  fails. Let  $S$  be “The KoF is so and so.” Russellians will call  $S$  false in worlds where France lacks a king. Strawsonians will say it is undefined. They agree, however, on  $S$ ’s truth-value in worlds where France has a unique king, and those are the only worlds that [1] cares about. Later I will be stipulating that  $S$ ’s “official” truth-value in a world goes with the truth-value of the IGNORE proposition, the one defined by [1].

[I]n a context where we both know that my neighbor is an adult male, I say, “My neighbor is a bachelor,” which, let us suppose, entails he is adult and a male. I might just as well have said “my neighbor is unmarried.” The same information would have been conveyed. (1999, 49)

The same information would indeed have been conveyed if by “information conveyed” we have in mind assertive content in the sense of [1] above, for (ignoring worlds where my neighbor fails to be an adult male), my neighbor is a bachelor if and only if he is unmarried.<sup>18</sup>

Never mind whether the IGNORE strategy can be attributed to Stalnaker; does it succeed in making  $\pi$  not a part of  $S$ 's asserted content but a determinant of that content? It does.  $\pi$  influences what  $S$  says by marking out the set of worlds on which  $S$  is defined. But  $\pi$  is not a part of what  $S$  says, for [1] makes  $S$  undefined in worlds where  $\pi$  is false, and it would be false in those worlds if  $S$  said in part that  $\pi$ .

The IGNORE proposition has some of the features we wanted. But what we mainly wanted was an  $S$  that could still be evaluated in worlds where  $\pi$  failed. And here [1] does not deliver at all. “The King of France is sitting in this chair” sounds to most people just false. But there is nothing in the IGNORE proposition to support this judgment, for the IGNORE proposition is undefined on worlds where France lacks a king.

Methodological digression: I said that “The KoF is sitting in this chair” sounds to most people just false. Why not go further and declare that it really *is* false? Strawson, for his part, is reluctant to take this further step. “The KoF is sitting in this chair” is *not* false in what he considers the term's primary sense: “*sometimes* [however] the word ‘false’ may acquire a *secondary* use, which collides with the primary one” (1954, 230).

One option is to follow Strawson in calling sentences like “The KoF is sitting in this chair” false only on a secondary use of that term, and sentences like “The US Chamber of Deputies has representatives from two major parties” true only on a secondary use of “true.” The task is then to explain why some gappy sentences *count* as false, while others count as true. Another option would be to follow Russell and call both of the above sentences false in the *primary* sense. The task would then be to explain why some primarily false sentences (“The man with the martini is a philosopher”) count as true, while others (“The King of France is bald”) count as neither true nor false.

Given that both theories (Russell's and Strawson's) need an analogous sort of supplement to deal with intuitive appearances of truth and falsity, either could

<sup>18</sup> Stalnaker would not identify what is said with a proposition defined only on  $\pi$ -worlds. Such an identification would make nonsense of passages like the following: “To make an assertion is to reduce the context set in a particular way . . . all of the possible situations incompatible with what is said are eliminated” (1999, 86). It is not clear to me how closely his notion of what is said—he sometimes calls it “the proposition expressed”—lines up with my assertive content, but certainly the correspondence is not exact.

serve as our jumping-off point; the choice is really between two styles of theoretical bookkeeping. That having been said, let's consider ourselves Strawsonians for purposes of this paper.  $S$ 's semantic content—what in context it *means*—will be a proposition defined only on  $\pi$ -worlds; it is semantic content that determines  $S$ 's truth-value.<sup>19</sup> Truth-value intuitions are driven not by what a sentence means, however, but by what it says: its asserted content.

So, “The KoF sits in this chair” strikes us as false because it says in part that someone sits in this chair. “The US Chamber of Deputies has representatives from two major parties” strikes us as true because it says that the House of Representatives has representatives from two major parties. Both of our remaining strategies are aimed at carving out a notion of asserted content that predicts truth-value intuitions in a way that semantic content is prevented from doing by the fact that it is undefined on worlds where  $\pi$  fails.

## 6. RESTORE

Let  $S$  be “The KoF is sitting in this chair.” Even if we agree with Strawson that  $S$  is lacking in truth-value, there is the feeling that it escapes on a technicality. The chair's emptiness is all set to falsify it, if France's lack of a king would just get out of the way. One response to this obstructionism is to say, fine, let's *give* France a king; then  $S$ 's deserved truth-value will shine through. This is the idea behind RESTORE. Instead of *ignoring* worlds where  $\pi$  fails, we attempt to *rehabilitate* them, in the sense of bringing them back into line with  $\pi$ . Of course one can't literally turn a non- $\pi$  world into a  $\pi$ -world, so in practice this means looking at  $S$ 's truth-value in the closest  $\pi$ -worlds to  $w$ .

Now, for  $S$  to be true (false) in the  $\pi$ -worlds closest to  $w$  is, on standard theories of conditionals, precisely what it takes for a conditional  $\pi \rightarrow S$  to be true (false) in  $w$ . So we can let the idea be this:

$$[2] \ S \text{ is true (false) in } w \text{ iff } \pi \rightarrow S \text{ is true (false) in } w.^{20}$$

Why does “The KoF is sitting in this chair” strike us as false? Even if France is supplied with a king, still he is not to be found in this chair. Why does “The KoF is bald” strike us as lacking in truth-value? Supplying France with a king leaves the issue still unresolved; in some closest worlds the added king is bald, in others not.<sup>21</sup> So the RESTORE strategy has *prima facie* a lot going for it.

<sup>19</sup> Von Fintel 2004 and Beaver and Kraemer (2001) also take this option. Because sentences and their semantic contents have the same truth-value (if any) in all worlds, we can be casual (sloppy) about the distinction between them. So, for instance, it makes no difference to an argument's validity whether we think of it as made up of (i) sentences, (ii) the propositions that are those sentences' semantic contents, or (iii) sentences and propositions combined.

<sup>20</sup> I assume that  $\pi \rightarrow S$  is false iff  $\pi \rightarrow \sim S$  is true.

<sup>21</sup> See Lasersohn 1993 and von Fintel 2004.

I don't doubt that for *some* similarity relation and *some* associated similarity-based conditional, [2] gives the right results. But if we confine ourselves to the conditionals we know best and have intuitions about—the indicative and the subjunctive—the strategy fails. Let me give some examples before attempting a diagnosis.

Bertrand Russell, invited to imagine what he could possibly say to God if his atheism proved incorrect, replied (not an exact quote), “I would ask him why he did not provide more evidence of his existence.” I infer from this that Russell accepted a certain indicative conditional

*G.* If God exists, he is doing a good job of hiding it.

Now the consequent of this conditional presupposes what its antecedent affirms; so *G* is of the form  $\pi \rightarrow S$ , read as *if it is the case that  $\pi$ , then it is the case that  $S$* . This according to [2-ind] is the condition under which what *S* says is true. But then it would seem that *S* ought to count for Russell as true, given that he accepts *G*. And something tells me that it does *not* strike Russell as true that God is doing a good job of hiding his existence.

So this remark of Russell's shows that [2] in its indicative version does not give a correct account of asserted content. Now consider a different Russell remark: “If there were a God, I think it very unlikely that he would have such an uneasy vanity as to be offended by those who doubt his existence.” From this it seems that Russell would have accepted

*H.* If there were a God, he would be generous to doubters.

*H* is of the form  $\pi \rightarrow S$ , read as *if it were the case that  $\pi$ , it would be the case that  $S$* . This according to [2-sub] is the condition under which what *S* says is true. So it would seem that *S* ought to count for Russell as true, given that he accepts *H*. But Russell is not at all inclined to think that God is generous to doubters.

Non-theological example: Is the King of France at this moment somewhere in the northern hemisphere? Of course not. But the corresponding conditionals are plausibly correct; that is where he would be, if he existed, and that is where he is, if he exists.

[2-sub] does get “The King of France is sitting in this chair” right, for the King, if he existed, would not be in this chair. But imagine for a moment that this chair is the long lost French throne; the King of France *would* (let's say) be sitting in this chair, if France had a king. [2-sub] predicts that our intuitions should shift. But it does not make it any more plausible to suppose that the King of France *is* sitting in this chair to be told that he *would* be sitting in it if France had a king. Imagine now that this chair is the long lost French throne and French kings, if any, are master illusionists; if France has a king, he *is* sitting in this chair. This does not affect our truth-value intuitions at all. It is enough for them that the chair is empty.

The problem we are finding with [2] (I will focus for simplicity on [2-sub]) is an instance of what used to be called the “conditional fallacy.” According to Shope (1978, 402—I have taken some liberties), the conditional fallacy is

A mistake one makes in analyzing a statement  $p$  by presenting its truth as dependent upon the truth of a conditional of the form: ‘If  $a$  were to occur, then  $b$  would occur’, when one has overlooked the fact that although statement  $p$  is actually true, if  $a$  were to occur, it would undermine  $p$  and so make  $b$  fail to occur.

Philosophers have tried, for instance, to analyze dispositions in counterfactual terms:

$x$  is fragile = if  $x$  were to be struck, it would shatter.

But  $x$  would not shatter if the molecular properties  $M$  making it fragile go away the moment that  $x$  is struck. What we meant to say, it seems, is that

$x$  is fragile = if  $x$  were struck and retained  $M$ , it would shatter.

[2-sub] tries to analyze false-seemingness in counterfactual terms:

$S$  counts as false = if  $\pi$ ,  $S$  would be false.<sup>22</sup>

But suppose  $S$  counts as false in virtue of certain facts  $F$ , and restoring  $S$ ’s presuppositions chases those facts away. (Europe would not have been King-of-France-free if France had had a king.) What we should have said, it seems, is that

$S$  counts as false = if  $\pi \& F$ ,  $S$  would be false

This is *essentially* what we do say in the next few sections. I mention this now because the motivation to be offered below is different, and we won’t be stopping to connect the dots.

## 7. SAY-MORE

A passage discussed earlier deserves a second look. Stalnaker had us choosing between “My neighbor is a bachelor” and “My neighbor is unmarried,” it being understood that my neighbor is an adult male. He says that the same information would be conveyed whichever sentence we chose. But in a part of the passage we didn’t get to, he puts the word “increment” before “information”: “the *increment of information*, or of content, conveyed by the first statement is the same as that conveyed by the second” (1999, 49).

<sup>22</sup> Perhaps the fallacy comes in an indicative version too. One is tempted to analyze “Jones is totally reliable” as: if Jones says  $X$ ,  $X$  is true. But if Jones says that  $0 = 1$ , that means not that  $0 = 1$  but that Jones is unreliable.

The word *increment* suggests that we are to ask what *more* it takes for  $S$  to be true, supposing the requirement of  $\pi$ 's truth is waived or assumed to be already met. This is the idea behind SAY-MORE. What  $S$  says, its assertive content, is identified with what *more*  $S$  asks of a world than that it should verify  $\pi$ .<sup>23</sup>

Determining these additional requirements may sound like a tricky business; but it is not so different from something we do every day, when we look for the missing premises in an enthymematic argument. To ask what further conditions (beyond  $\pi$ ) a world has to meet to be  $S$  is essentially to ask what premises should be added to  $\pi$  to obtain a valid argument for  $S$ :

$$\frac{\pi}{\frac{???}{S}}$$

So we can put the SAY-MORE strategy like this:

[3]  $S$  is whatever bridges the logical gap between  $\pi$  and  $S$ .

Of course, the gap might be bridgeable in more than one way. I propose to finesse this issue for now by letting  $S$  be the result of lumping all otherwise qualified gap-bridgers together. So, for instance,

France has exactly one king.  
 $\frac{???}{\text{The King of France is sitting in this chair.}}$

becomes valid if for ??? we put either "Some French king is sitting in this chair" or "All French kings are sitting in this chair." Assuming both statements bridge the gap equally well (see below), the assertive content is "Some and all French kings are sitting in this chair."

A lot more needs to be said, obviously, and some of it will be said in the next section. Right now, though, I want to try [3] out on a series of examples, one from Strawson, two adapted from Strawson, one from Donnellan, one from Kripke, and one from Langendoen.

- A The lodger next door offered me twice that sum.
- B The author of *Principia Mathematica* also wrote *Principia Ethica*.
- C All ten solar planets are inhabited.
- D The man drinking a martini is a philosopher.
- E Jones is burning the leaves.
- F My cousin is not a boy anymore.

<sup>23</sup> Suppose we use  $\text{prop}(\pi)$  for the properties a world must have to verify  $\pi$ ,  $\text{prop}(S)$  for the properties needed to verify  $S$ ; and  $\text{prop}(S|\pi)$  for the *additional* properties  $\pi$ -worlds must have to verify  $S$ . Then it is not in general the case that  $\text{prop}(S|\pi) = \text{prop}(S) - \text{prop}(\pi)$ . An analogy might be this. The rich can get into heaven only by having the property  $S$  of giving lots of money to charity.  $S$  is thus in  $\text{prop}(H|R)$ . But  $S$  is not in  $\text{prop}(H) - \text{prop}(R)$ , or even in  $\text{prop}(H)$ ; the deserving poor get into heaven without it.

All six sentences are meant to strike us as false—the first because there is no lodger next door; the second because neither *PM* author wrote *PE*; the third because most solar planets (they number nine, not ten) are uninhabited; the fourth because Daniels (who is in fact drinking water) is not a philosopher but an engineer; the fifth because that man (it's really Smith) is not burning leaves but raking them; and the sixth because my cousin (whether a boy or not) is only eight years old.

How would Stalnaker explain the appearance of falsity in these cases? This is to some extent speculative, but here is what I suspect he would say. *A* seems false for Russellian reasons: it is equivalent to a conjunction one of whose conjuncts is “There is a lodger next door.” *B* seems false for supervaluational reasons: it is false on all admissible disambiguations. *C* and *E* seem false for the sort of reason Strawson offered (section 4): we amend them to “All nine solar planets are inhabited” and “Smith is burning the leaves” before assigning a truth-value. *D* seems false for Kaplanian reasons: its character applied to the context of utterance issues in a falsehood, viz. Daniels is a philosopher. *F* seems false because I use it to make an assertion not about my cousin's sex (that's presupposed) but my cousin's age.

The hope is that we can replace these various explanations with one, perhaps closest in spirit to Stalnaker's proposal about *F*: the sentences seem false because what they assert is false. [3] tells us how to find the propositions asserted; we ask what assumptions have to be added to

- $\pi_A$  There is exactly one lodger next door.
- $\pi_B$  *Principia Mathematica* had exactly one author.
- $\pi_C$  There are exactly ten solar planets.
- $\pi_D$  That man [pointing] is the man drinking a martini.
- $\pi_E$  That man [pointing] is Jones.
- $\pi_F$  My cousin is a male human being.

for it to follow that

- A* The lodger next door offered me twice that sum.
- B* The author of *Principia Mathematica* also wrote *Principia Ethica*.
- C* All ten solar planets are inhabited.
- D* The man drinking a martini is a philosopher.
- E* Jones is burning the leaves.
- F* My cousin is not a boy anymore.

The needed assumptions would seem to be

- A* Some and all lodgers next door offered me twice that sum.
- B* Some and all *Principia Mathematica* authors also wrote *Principia Ethica*.

*C* All solar planets are inhabited.

*D* That man is a philosopher.

*E* That man is burning the leaves.

*F* My cousin is an adult.

*A–F*, the asserted contents of *A–F*, really are what *A–F* only appear to be, namely false.<sup>24</sup> The suggestion (once again) is that this is not a coincidence. *A–F* appear false because of the genuine falsity of what they assert or say.

I have been stressing the role asserted content plays in explaining felt truth-value, but it is also relevant to judgments about what is said, contributing in this second way even where truth-value intuitions are lacking. This is the application that matters to Stalnaker:

it is possible for . . . presuppositions to vary from context to context, or with changes in stress or shifts in word order, without those changes requiring variation in the semantic interpretation of what is said. This should make possible a simpler semantic theory . . . (1999, 53)

There is that much less need to multiply meanings if “one [can] use the same sentence” against the background of different assumptions to assert different things. (Grice of course makes similar claims on behalf of implicature.<sup>25</sup>) [3] shows why asserted content would fluctuate in this way: the shape of the logical gap between  $\pi$  and *S* is clearly going to depend in part on  $\pi$ . To illustrate with the Donnellan case, the gap between  $\pi_D$  and *D* is filled by a proposition about Daniels because that is who we presume to be drinking a martini; if we decide it is really O’Leary, then the gap-filler changes accordingly. Or consider Stalnaker’s elaboration of Langendoen’s example:

normally, if one said “my cousin isn’t a boy anymore” he would be asserting that his cousin had grown up, presupposing that he is male. But one might, in a less common context, use the same sentence to assert that one’s cousin had changed sexes, presupposing that she is young. (1999, 53–4)

The first proposition he mentions (my cousin has grown up) corresponds to the missing premise in

My cousin is and always has been a male human being.

???

My cousin is not a boy any more.

<sup>24</sup> The term “asserted content” might be in some cases misleading, since one does not hear “The lodger next door offered me twice that sum” as *asserting* that some and all lodgers next door offered me twice that sum. Other terms sometimes used are “allegational,” “proffered,” or “at-issue” content.

<sup>25</sup> See “Further Notes on Logic and Conversation” in Grice (1989), esp. the discussion of “modified Occam’s razor” on 47–9.

while the second (my cousin has changed sexes) is the premise one needs to make a valid argument of

My cousin is and has always been a human child.  
 ???  
 -----  
 My cousin is not a boy any more.

A final example concerns the cognitive content of proper names. How are we to reconcile Mill's idea that names mean their referents with Frege's observation that a name's contribution to cognitive content is not predictable from its meaning so defined? Stalnaker's candidate for the role of (unpredicted) cognitive content is the diagonal proposition, but assertive content can be helpful in this regard as well. Suppose that "*n*" is a name and that *n* is presumed to be the so and so. Then "*n* is F" asserts that the so and so is F (that is what it takes to get from the stated presumption to the conclusion that *n* is F). This is why "A meteor is about to destroy the Earth" is experienced as saying that a meteor is about to destroy this very planet.

## 8. DEFINITIONS

There is more to bridging the gap between  $\pi$  and  $S$  than combining with  $\pi$  to imply  $S$ . It is crucial that  $S$  not be a "bridge too far" at either end. Let me explain what I mean by that.

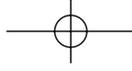
$X$  is a bridge too far at the  $S$  end if it combines with  $\pi$  to imply *more* than  $S$ . So, to take again the Donnellan example, the proposition we want is *That man is a philosopher*, not *That man is a philosopher & snow is white*. The latter proposition combines with  $\pi$  to yield "The man with the martini is a philosopher & snow is white," which is a stronger conclusion than we were aiming for.

$X$  would be a bridge too far at the  $\pi$  end if it made for redundancy in the premises—if it repeated material already present in  $\pi$ . The proposition we want is *That man is a philosopher*, not *That man is a martini-drinking philosopher*. It is stated already in  $\pi$  that he is drinking a martini, and there is no need to repeat what has already been stated.

How do we enforce the requirement of not being a bridge too far at the  $S$  end? Suppose  $X$  and  $\pi$  imply a stronger statement than  $S$  = a statement that  $S$  does not imply. Then  $S$  does not imply  $X \& \pi$  (or it would imply the stronger statement). Turning this around, if we stipulate that  $S$  *does* imply  $X \& \pi$ , that will prevent  $X$  from being a bridge too far at the  $S$  end.<sup>26</sup>

How do we enforce the requirement of not being a bridge too far at the  $\pi$  end? Suppose we have our hands on the reason why  $X$  is true (false)—its truth-maker

<sup>26</sup> There is no question of  $S$  not implying  $\pi$ —we have stipulated that  $S$  has truth-value only if  $\pi$  is true—so the requirement is really just that  $S$  should imply  $X$ .



or falsity-maker.  $X$  has a trace of  $\pi$  in it iff  $X$  could not be true (false) *for that reason* unless  $\pi$  too were true (false). So

- [4]  $X$  is  $\pi$ -free iff  $X$  is true (false) and could be true (false) for the same reason, that is, with the same truth-(falsity-)maker, even if  $\pi$  were false (true).

I will leave to an appendix my attempt at a theory of truth- and falsity-makers. But the idea is this. Truth-preservation across all worlds  $w$  can be called *global* implication.<sup>27</sup> *Local* implication (in some particular  $w$ ) is truth-preservation across all situations in  $w$ . A fact in  $w$  is a proposition true in  $w$ .

- [5] A truth-maker for  $X$  in  $w$  is a fact that implies  $X$  globally and is implied by  $X$  locally.<sup>28</sup>

Now we can explain what is involved in bridging the gap between  $\pi$  and  $S$ . Certainly  $X$  should combine with  $\pi$  to imply  $S$ . Also, though  $S$  should return the favor; it should imply  $X \& \pi$ . So much is basically to say that although  $X$  may well be defined on additional worlds,  $X$  restricted to the  $\pi$ -worlds is true/false in the very same worlds as  $S$ . Since  $X$  extends  $S$  beyond the  $\pi$ -worlds, let's call it an *extension* of  $S$ . Finally  $X$  should be  $\pi$ -free. All in all, then,

- [6]  $X$  bridges the gap between  $\pi$  and  $S$  iff  $X$  is a  $\pi$ -free extension of  $S$ .

Asserted content was to be the sum or conjunction of gap-bridgers, so

- [7]  $S$  is the conjunction of  $S$ 's  $\pi$ -free extensions.

This can be simplified, however. Extensions are just maximum-strength implications; they are implications false in as many  $\pi$ -worlds as possible, compatibly with still being implied by  $S$ . Like any maximum-strength implications, they are conjunctions of regular implications. But then conjoining extensions is conjoining conjunctions of implications, which by the associative law for conjunction is just conjoining implications. So we obtain the same definition as [7] in a more digestible form if we say instead that

<sup>27</sup> Implication should preserve definedness as well as truth. The reason is this. "The KoF is bald" should not  $\pi$ -free imply the falsehood, "Among the bald people is a French king." The latter is false because there are no French kings among the bald people, which is compatible with France's having a unique king. This *would* be a case of  $\pi$ -free implication, if it were a case of implication. But it is not a case of implication, because although truth is preserved, definedness is not. (There could be a world lacking in bald people where France had a unique king.) On an intuitive level, requiring  $X$  to be defined wherever  $S$  is defined is requiring that  $X$ 's presuppositions be no stronger than  $S$ 's. This fits with the idea that  $S$  counts as false because it implies something whose weaker presuppositions allow it to be false where  $S$  is undefined.

<sup>28</sup> Falsity-makers are similar. It is not assumed that  $X$  has only one truth- or falsity-maker. A disjunction might be made true either via its left disjunct or its right. See Appendix.



[8]  $S$  is the conjunction of  $S$ 's  $\pi$ -free implications.<sup>29</sup>

This way of putting it further clarifies why an  $S$  that is undefined due to presupposition failure might nevertheless strike us as false.  $S$  does not seem false merely because it implies a falsehood, for all the sentences we are talking about do that much, just by virtue of implying  $\pi$ . (e.g., “The KoF is bald” implies that France has a unique king.)  $S$  seems false because it implies a falsehood, the reasons for whose truth-value have nothing to do with  $\pi$ .

“The KoF is sitting in this chair” implies “Someone is sitting in this chair.” “Someone is sitting in this chair” doesn't suffer from presupposition failure, so we can evaluate it; it is false. Moreover, and this is crucial, “Someone is sitting in this chair” is false for a reason that could still obtain even if France had a unique king, viz. that the chair is empty.

Compare “The KoF is bald.” It implies “France has a king.” “France has a king” doesn't suffer from presupposition failure, so we can evaluate it; it is false. So “The KoF is bald” counts as false, if “France has a king” is  $\pi$ -free. Could “France has a bald king” have been false *for the same reason* even if France had a king? No, it could not. The reason “France has a bald king” is false is that *France has no king*. Clearly it could not have been false *for that reason* in a world where  $\pi$  was true—a world where France had a unique king.

caps

## 9. CLAIMS

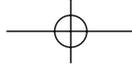
A theory of NCPF should address itself to three questions. First, how is it possible, that is, how can presupposition failure ever fail to be catastrophic? Second, why does catastrophe strike in some cases but not others? Third, limiting ourselves to the “good” cases, why do some of these sentences strike us as true and others as false?

Stalnaker gave the outlines of an answer to our first question: how is NCPF so much as possible?  $\pi$  helps to determine the proposition expressed;  $\pi$ 's truth-value is not directly relevant to its role as proposition-determiner; and the proposition determined is not limited to worlds where  $\pi$  is true. I call this the outline of an answer because Stalnaker doesn't really explain how false  $\pi$ s *can* play the same determinative role as true ones.<sup>30</sup>

Now that we have some idea of the mechanism by which  $\pi$  exerts its influence, we can see why truth-value doesn't come into it. Those of us who use stacks of books as bookends know that false books perform just as well in this capacity

<sup>29</sup> This should be understood to mean “natural, intelligible implications” (and above, “natural, intelligible extensions”). Otherwise asserted content becomes hard to distinguish from semantic content, as the intersection of *all*  $\pi$ -free propositions implied is liable to be defined only on the  $\pi$ -worlds and the actual world.

<sup>30</sup> Leaving aside special cases like the referential use of definite descriptions.



as true ones. It's the same with presuppositions. The shape of the logical gap between  $\pi$  and  $S$  is defined by  $\pi$ 's content; whether that content obtains doesn't much matter.<sup>31</sup>

So that's our explanation of how NCPF is possible. It may seem that we have succeeded too well. If  $\pi$ 's falsity is irrelevant to its role in determining asserted content, why is presupposition failure ever catastrophic?

This would be a good question if catastrophic presupposition failure had been characterized as presupposition failure resulting in the loss of *asserted content*. But that is not how we explained it. Presupposition failure is catastrophic, we said, if it has the result that  $S$  makes no *claim*. And having an asserted content does not suffice for making a claim.

The reason is this. Part of what is involved in  $S$ 's making a claim is that for matters to stand as  $S$  says is for them NOT to stand as  $\sim S$  says. It can happen that  $S$  is so tainted by its association with  $\pi$  that  $S$  and  $\sim S$  cease to disagree when  $\pi$  fails. A sentence whose negation is (counts as) just as true as itself takes no risks and cannot be used to convey real information.

- [9]  $S$  makes a claim iff: if  $S$  is true, then  $\sim S$  is false.  
 $S$  makes no claim:  $S$  is true and  $\sim S$  is also true

Now we can define the central notion of this paper:

- [10]  $S$  is a case of *catastrophic presupposition failure* iff  $\pi$ 's falsity has the result that  $S$  makes no claim.

The poster child here is "The KoF is bald." To find its assertive content we ask, what beyond France's having a unique king is required for the KoF to be bald? The candidates are, let's suppose, (i) *France has a bald king*, and (ii) *any French kings are bald*. We have already seen that (i) is not  $\pi$ -free, because it is false for a reason incompatible with  $\pi$ , namely France's lack of a king. What about (ii)? Its truth-maker (again, France's lack of a king) is consistent with  $\pi$  being false, so (ii) is  $\pi$ -free. It appears that there is nothing for "The KoF is bald" to say but *Any French kings are bald*, and nothing for its negation to say but *Any French kings are non-bald*, that is, *France has no bald kings*. Both propositions are true; so according to [9], "The KoF is bald" makes no claim, and according to [10], "The KoF is bald" suffers from catastrophic presupposition failure.

## 10. "TRUE" AND "FALSE"

This leaves the question of what distinguishes the presuppositionally challenged sentences that count as true from the ones that count as false. First a stipulation:

<sup>31</sup> It can matter a little. If  $\pi$  is true, then any false implication  $X$  of  $S$  is automatically  $\pi$ -free;  $X$ 's falsity-maker coexists with  $\pi$  in this world, so obviously the two are compatible. Likewise if  $\pi$  is false, then  $S$ 's true implications are automatically  $\pi$ -free.



the claim  $S$  makes—when it makes a claim—is its asserted content. A sentence’s felt truth-value goes with the truth-value of the claim it makes:

[11]  $S$  counts as true (false) iff it makes a true (false) claim.

Take “The KoF is sitting in this chair.” What  $\pi$ -free implications does it have? One obvious implication is *The chair contains a French king*. This is false because *No one is sitting in the chair*, or perhaps because *No king is . . .*, or *No French king is . . .*<sup>32</sup> All of these are compatible with France’s having a unique king. It looks, then, like “The KoF is sitting in this chair” claims in part that *The chair contains a French king*. That claim is false, so the sentence that makes the claim (“The KoF is sitting in this chair”) counts as false.

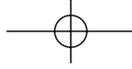
A couple of examples, finally, of counting as *true* despite presupposition failure. The first will use a concrete term, the second, to move us back to the ontological relevance of NCPF, will use a term that’s abstract.

A long time ago there were two popes, one in Rome, the other in Avignon. Imagine that we check into a monastery one fine night, and scrawled on the bed-post we read, “The pope slept here.” A bit of research reveals that both popes in fact slept in the bed in question. Then it seems to me that the inscription strikes us as true; for although it was making a stronger claim than it knew, this stronger claim was correct. To check this against the theory, we must hunt around for  $\pi$ -free implications of “The pope slept here.” One such implication is that *some* pope slept here; another is that *all* popes slept here. So far, then, it looks like the assertive content is that some and all popes slept here. It is because this is true in the imagined circumstance that “The pope slept here” counts for us as true.

Looking into my back yard last Monday, I saw one cat; looking into the yard last Tuesday, I saw two cats; and so on. I ~~now form the hypothesis~~ hypothesize that this pattern will continue forever<sup>33</sup>—to state the hypothesis more explicitly,

<sup>32</sup> Note that *France lacks a king* does not make “This chair contains a French king” false, since it does not imply the latter’s presupposition that there is such a thing as this chair. What about *France lacks a king and there is such a thing as this chair*? This is formally eligible but a better—more proportional—candidate for the role is *This chair is empty* (see Yablo 2003 and the Appendix). Objection: If “The KoF sits in this chair”  $\pi$ -free implies the false “A French king sits in this chair,” shouldn’t “The KoF is heavier than this chair”  $\pi$ -free imply the false “A French king is heavier than this chair”? Yet “The KoF weighs more than this chair” does *not* strike us as false. I reply that “A French king is heavier than this chair” is not  $\pi$ -free, because in *this* case there is no better candidate for falsity-maker than *France lacks a king and there is such a thing as this chair*. No simple fact about the chair itself falsifies “A French king is heavier than this chair” as the chair’s emptiness falsifies “A French king sits in this chair.” More generally, “The KoF bears  $R$  to  $x$ ” does not strike us as false when  $R$  is an “internal relation” like taller-than or heavier-than—a relation that obtains in virtue of intrinsic properties of the relata. The proposed explanation is that facts purely about  $x$  do not suffice for the falsity of “A French king bears  $R$  to  $x$ ”; France’s lack of a king has to be brought in, which makes the implication no longer  $\pi$ -free. (See in this connection Donnellan 1981 and von Fintel 2004.)

<sup>33</sup> The example is from Burgess and Rosen (2005).



“For all  $n$ , the number of cats in my yard on the  $n$ th day =  $n$ .”

I presuppose in saying this that no matter what day it is, there is a unique thing that numbers the cats in my yard on that day (and more generally that whenever there are finitely many Fs, there is a unique thing that numbers them). Now consider the statements on this list:

- on the first day, there is one cat in my yard
- on the second day, there are two cats in my yard
- on the third day, there are three cats etc.
- etc.

All of them are implied by my hypothesis, and it is easy to see that each is a  $\pi$ -free implication. A typical falsity-maker is the fact that on the third day, there are no cats in my yard. That there are no cats in my yard can happen just as easily in platonistic worlds (where  $\pi$  holds) as in wholly concrete worlds. A typical truth-maker is the fact that on the third day, the cats in my yard are Zora, Teasel, and Yossele. For the cats in my yard to be Zora, Teasel, and Yossele can happen just as easily in concrete worlds (where  $\pi$  fails) as in platonistic worlds.

Now I haven't argued that these are *all* the  $\pi$ -free implications. But if they are, then what my hypothesis *says* is that on the first day there is one cat, on the second there are two, and so on. This fits with our intuitive sense that the hypothesis counts as true or false according to how many cats my yard contains on which days; the existence of numbers plays no role whatsoever. Note the analogy with the King of France. Just as “The KoF is sitting in this chair” counts as false because of the chair's material contents—nothing to do with French royalty—“The number of cats on the  $n$ th day =  $n$ ” counts as true, if it does, because of my yard's material contents—nothing to do with numbers.

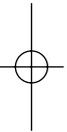
### 11. PARTING THOUGHTS ON ABSTRACT ONTOLOGY

Nominalists maintain that abstract terms do not refer. They will find it suggestive, then, that (supposedly empty) abstract terms make in some cases the same sort of contribution to felt truth-value as definitely empty concrete terms do.

Platonists will complain that empty concrete terms make a *negative* contribution; simple King-of-France sentences almost all count as false, to the extent that they make a claim at all. One would expect the emptiness of abstract terms, if they were empty, to manifest itself the same way. But the sentences we construct with abstract terms very often strike us as true.

I agree that the failure of a *concrete* term to refer prevents it from exercising positive semantic influence. But there is a reason for this. “The King of France”'s semantic contribution goes way beyond our notions of what a French king would have to be like. He is (or would be, if he existed) an original source of information

↗  
↘



of the type that makes simple King-of-France sentences count as true. Numbers, by contrast, are not (would not be) an original source of information on any topic of interest; their contribution is exhausted by what they are *supposed* to be like. This makes the presupposition that numbers exist “fail-safe” in the sense that its failure makes (or would make) no difference whatever to which applied arithmetical sentences count as true. I am tempted to conclude that nothing in the felt truth-values of those sentences has any bearing on the issue of whether numbers exist.

## APPENDIX

Situations are parts of worlds; worlds are maximal situations; truth-in-a-world is a special case of truth-in-a-situation. Suppose that  $A$  is a sentence and situation  $s$  is part of world  $w$ . Then  $s$  verifies (falsifies)  $A$  only if it contains everything potentially relevant to  $A$ 's truth-value in  $w$ . Thus for  $s$  to verify “All swans are white,” it is not enough that all the swans in  $s$  are white;  $s$  should also contain all of  $w$ 's swans. The formal upshot is a condition called *persistence*: if  $A$  is true (false) in  $s$  and  $s$  is part of  $s^*$ , then  $A$  is true (false) in  $s^*$ .<sup>34</sup> Now we introduce two notions of implication (the first is familiar, the second not):

$A$  implies  $B$  globally iff for all worlds  $w$ ,  $A$  is true in  $w$  only if  $B$  is true in  $w$ .<sup>35</sup>

$A$  implies  $B$  locally in  $w$  iff for all  $s \leq w$ ,  $A$  is true in  $s$  only if  $B$  is true in  $s$ .<sup>36</sup>

Here is our first stab at a definition of truth-maker. A truth-maker for  $X$  in  $w$  is a  $T$  that implies  $X$  across all possible worlds, and that holds in every  $w$ -situation where  $X$  holds:

$T$  makes  $X$  true in  $w$  iff

- (a)  $T$  is true in  $w$ ,
- (b)  $T$  implies  $X$  globally,
- (c)  $X$  implies  $T$  locally in  $w$ .

( $F$  makes  $X$  false in  $w$  iff it makes  $\sim X$  true.) This runs into a problem, however.<sup>37</sup>

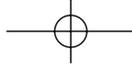
It could happen that  $X$ 's truth in  $w$  is overdetermined by  $T_1$  and  $T_2$  (e.g.,  $X$  might be a disjunction of unrelated disjuncts both of which are true). Then  $X$  may well lack a truth-maker in the above sense. It won't imply  $T_1$  in  $w$  because there are  $w$ -situations where  $X$  holds thanks instead to  $T_2$ , and it won't imply  $T_2$  because there are  $w$ -situations where  $X$  holds thanks instead to  $T_1$ . Still, there ought to be a  $v_1 \leq w$  such that  $X$  implies  $T_1$  in  $v_1$  and a  $v_2 \leq w$  such that

<sup>34</sup> See Kratzer for an enlightening discussion of persistence.

<sup>35</sup> I ignore that implication should preserve definedness as well. See n. 27.

<sup>36</sup> Compare the definition of lumping in Kratzer (1989).

<sup>37</sup> Originally raised by Heim against Kratzer.



$X$  implies  $T_2$  in  $v_2$ . So rather than asking  $X$  to imply  $T$  in  $w$ , we should ask it to imply  $T$  in some subsituation of  $w$ .

$T$  makes  $X$  true in  $w$  iff for some  $v \leq w$

- (a)  $T$  is true in  $v$ ,
- (b)  $T$  implies  $X$ ,
- (c)  $X$  implies  $T$  in  $v$ .

Now that we are explicitly contemplating multiple truth- and falsity-makers, we need to adjust the definition of  $\pi$ -freedom:

$X$  is  $\pi$ -free in  $w$  iff  $X$  is true in  $w$  and it has a truth-maker that holds also in worlds where  $\pi$  is false, or  $X$  is false in  $w$  and it has a falsity-maker that holds also in worlds where  $\pi$  is true.

This runs into a different problem. “France has a bald king” had better not be a  $\pi$ -free consequence of “The King of France is bald,” or the latter will count as false, which it shouldn’t. The obvious falsity-maker is *France has no king*, which is indeed not compatible with  $\pi$ . But another technically eligible falsity-maker is *France has no bald kings*. This is compatible with France’s having a unique king, and since our definition requires only that *some* falsity-maker be transportable to  $\pi$ -worlds, it seems we are sunk. The solution I suggest is to impose a proportionality requirement along roughly the lines of Yablo 2003. *France has no bald kings* is needlessly complicated in that a strictly simpler condition (*France lacks a king*) still satisfies conditions (a)–(c). Thus we should think of (a)–(c) as defining *candidacy* for the role of truth- (falsity-) maker, and add that the successful candidate should not involve gratuitous complications in whose absence (a)–(c) would still be satisfied.

## REFERENCES

- Beaney, Michael. 1997. *The Frege Reader* (Oxford: Blackwell)
- Beaver, David. 2001. *Presupposition and Assertion in Dynamic Semantics* (Palo Alto: CSLI Press)
- Beaver, David, and Emil Krahmer. 2001. “A partial account of presupposition projection.” *Journal of Logic, Language and Information* 10(2): 147–82
- Bezuidenhout, A. and Reimer, M. 2004. *Descriptions and Beyond* (Oxford: OUP)
- Burgess, John, and Gideon Rosen. 2005. “Nominalism Reconsidered,” to appear in Stewart Shapiro (ed.), *Handbook of Philosophy of Mathematics and Logic*
- Burton-Roberts, Noel. 1989. *The Limits to Debate: A Revised Theory of Semantic Presupposition* (Cambridge: Cambridge University Press)
- Cohen, Ariel. 2000. “The King of France Is, In Fact, Bald,” *Natural Language Semantics* 8: 291–5
- Cole, P. (ed.). 1981. *Radical Pragmatics* (New York: Academic Press)



- Donnellan, Keith. 1966. "Reference and Definite Descriptions," *Philosophical Review* 75: 281–304
- 1981. "Intuitions and Presuppositions," in Cole (1981, 129–42)
- von Fintel, Kai. 2008. "What is Presupposition accommodation, again?," *Philosophical Perspectives* 22(1): 137–70
- 2004. "Would You Believe It? The King of France is Back," in Bezuidenhout and Reimer 2004, 315–41
- Frege, Gottlob. 1872. "On Sense and Meaning," page references to Beaney 1997
- 1918. "The Thought," page references to Beaney 1997
- Glanzberg, Michael. 2005. "Presuppositions, truth values, and expressing propositions," In G. Preyer and G. Peter (eds.), *Contextualism in Philosophy* (Oxford: OUP)
- 2002. "Context and Discourse," *Mind and Language* 17: 333–75
- Grice, H. Paul. 1989. *Studies in the Way of Words* (Cambridge, Mass.: Harvard University Press)
- Heim, Irene. 1983. "On the Projection Problem for Presuppositions," *WCCFL* 2
- Humberstone, Lloyd. 2000. "Parts and Partitions," *Theoria* 66: 41–82
- Kratzer, Angelika. 1989. "An Investigation into the Lumps of Thought," *Linguistics and Philosophy* 12: 607–53
- Kripke, Saul. 1977. "Speaker's Reference and Semantic Reference," originally in *Midwest Studies in Philosophy* II, as reprinted in French, Uehling, and Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language* (University of Minnesota Press: Minneapolis), 6–27
- Laserson, Peter. 1993. "Existence Presuppositions and Background Knowledge," *Journal of Semantics* 10: 113–22
- van der Sandt, Rob. 1989. "Anaphora and Accommodation," in Renate Bartsch, Johann van Benthem and Peter van Emde Boas (eds.), *Semantics and Contextual Expression* (Foris: Dordrecht)
- Shanon, Benny. 1976. "On the Two Kinds of Presupposition in Natural Language," *Foundations of Language* 14: 247–9
- Shope, Robert. 1978. "The Conditional Fallacy in Contemporary Philosophy," *Journal of Philosophy* 75: 397–413
- Simons, Mandy. 2003. "Presupposition and Accommodation: Understanding the Stalnakerian Picture," *Philosophical Studies* 112: 251–78
- Stalnaker, Robert. 1970. "Pragmatics," *Synthese* 22, page references to the reprint in Stalnaker 1999
- 1974. "Pragmatic Presuppositions," in Munitz and Unger (eds.), *Semantics and Philosophy* (New York: NYU Press), page references to the reprint in Stalnaker 1999.
- 1978. "Assertion," in Cole (ed.), *Syntax and Semantics*, vol. 9, page references to the reprint in Stalnaker 1999
- 1999. *Content and Context* (Oxford: OUP)
- 2002. "Common Ground," *Linguistics and Philosophy* 25: 701–21
- Strawson, P. F. 1950, "On Referring," *Mind* 59, page references to the reprint in Strawson 1971
- 1954. "A Reply to Mr. Sellars," *Philosophical Review* 63: 216–31



- 1964, "Identifying Reference and Truth-Values," *Theoria* 30; page references to the reprint in Strawson 1971
- 1971. *Logico-Linguistic Papers* (London: Methuen)
- Yablo, Stephen. 2003. "Causal Relevance," *Philosophical Issues* 13: 316–29
- Zeevat, Hank. 1992. "Presupposition and Accommodation in Update Semantics," *Journal of Semantics* 9: 379–412



## Must Existence-Questions Have Answers?

*Are you lost daddy I asked tenderly.  
Shut up he explained.*

Ring Lardner, *The Young Immigrants*<sup>1</sup>

### 1. INTRODUCTION

I suppose, to go by the analogy with ethics, that first-order or “normative” ontologists debate what really exists, while second-order or meta-ontologists ponder those first-order debates. This paper concerns itself with two meta-ontological questions, one theoretical and one practical. The practical question, for a given type of entity  $X$ , is

FUTILITY: are debates about the existence of  $X$ s as futile and pointless as they can sometimes seem?

The theoretical question is

VACUITY: is anything genuinely at issue in debates about the existence of  $X$ s? is there a fact of the matter to be right or wrong about?

FUTILITY and VACUITY are related in that one reason for a debate to be futile and pointless is that there is nothing at issue in that debate; and there being nothing at issue in a debate would tend to vindicate the feeling that the debate is futile and pointless. But they are also to some degree independent, since a non-vacuous debate can still be futile and pointless, due, say, to a lack of evidence either way, or the fact that some views are so comprehensively misguided that any criticism could be rejected as question-begging.

Thanks to Karen Bennett for responding to an earlier version of this material at a symposium on ‘Metaontology and Existence’ at the 2004 meetings of the American Philosophical Association (APA). Thanks to Ted Sider, John Hawthorne, David Chalmers, David Papineau, Guy Rohrbaugh, and others in audiences at the 2004 Eastern APA, the 2005 Metametaphysics Conference at Australian National University, University College London, Syracuse University, and Auburn University. Eli Hirsch and Jonathan Schaffer gave extremely helpful written comments.

<sup>1</sup> I learned “‘Shut up,’ he explained” from Paul Boghossian. Thanks to Amanda Hale for pointing out I had got Lardner’s title wrong.

On the practical question, philosophers divide into two rough camps. One camp wants the debate about *X*s to continue, because they think it productive and take an interest in how it comes out. I will call this the first-order *ontologist's* camp.<sup>2</sup> Reasons for interest in the debate may differ. Some ontologists are genuinely uncertain whether *X*s exist and are hoping for enlightenment on the matter. More common, though, are ontologists who “know” how the debate comes out. Those who expect a positive verdict will be called *platonists* and those expecting a negative verdict will be *nominalists*.<sup>3</sup>

That's the first camp, then: the ontologists, of whom some are platonists and others are nominalists. A second camp takes a more quizzical perspective. Quizzicalists, as I'll call them, find it hard to take (some? all?) ontological debate seriously and hold out little hope for a successful resolution. Their advice to the engaged ontologist is to disengage and find something useful to do. Of course, quizzicalists will try to wrap that advice in hopefully edifying commentary about the ontologist's predicament; they will point out features of the dialectical situation that the ontologist may be overlooking. The advice is still in effect to shut up. But the quizzicalist's line is not, “‘Shut up,’ he *angrily demanded*,” but, like the daddy in the Lardner dialogue, “‘Shut up,’ he *explained*.”

## 2. GOALS AND DESIDERATA

This paper features four main characters: the ontologist O; two particularly opinionated types of ontologist, the platonist P and the nominalist N; and the quizzicalist Q. Q is the relatively neglected party here, so we will let him speak first.

Q. I adore the first-order ontologist; it's a stage I've been through myself. But his wide-eyed innocence about existence is beginning to wear me down. Perhaps it's because beneath the questions I sense an undercurrent of reproach. “Ignore me if you like,” he seems to be saying, “but then we'll just keep driving in circles.” Like the daddy in the dialogue, I resent the reproach and see no need to pull over and ask for directions.

O. If it's the daddy you're identifying with, then let's not forget that the daddy is lost. He really *should* ask for directions; he just won't admit it. Could it be that you, Q, are similarly in denial? If some objects exist and others don't, that would seem to be a fact of supreme metaphysical importance. What possible defense could there be for ignoring it?

<sup>2</sup> Sometimes omitting the “first-order”, on the theory that meta-ontology is not a branch of ontology.

<sup>3</sup> Consider this stipulative. The terms “platonist” and “nominalist” are often reserved just for the case where the *X*s are abstract objects, and that is the case I have mainly in mind too. But it will be useful later on to apply the terms more broadly, to disputes about the existence of (non-abstract items like) Lewisian worlds and arbitrary mereological sums.

Q. This takes us from FUTILITY, the practical question, to VACUITY, the theoretical one. The best defense against the charge of ignoring a fact is to deny that there is a fact there to ignore. Debates about the existence of numbers or . . . (how to continue this list is an important question, to which we return) are just empty; there is nothing at issue in them.

O. How am I supposed to even make sense of that claim? I can make sense of there being no fact of the matter as to whether someone is short (Tom Cruise, say), because there's a story to be told about how this situation arises. "Short" is a vague predicate; vague predicates have borderline cases;<sup>4</sup> and Cruise is a borderline case. (He is taller than definitely short people and shorter than people who are definitely not short.) Of course, there are various theories of what borderline-case-hood amounts to. Maybe "short" doesn't pick out any definite property; some candidates for the role of shortness Cruise has, others he lacks. Maybe it picks out different definite properties at different times or in different settings; he has some of these properties but not others. Maybe the word picks out the same definite property all the time, but it's a property that itself has borderline cases. These theories are different, but they all allow us to point to Cruise's borderline-case-hood as the reason there is no fact of the matter as to whether he is short.

O (continuing). Suppose we now ask how there can fail to be a fact of the matter about, say, the existence of the empty set. To go by our discussion above, the explanation should be this: "exists" is a vague predicate and  $\emptyset$  is a borderline case. But it is harder to make sense of a borderline case of existence than of shortness. It helped us understand Cruise's borderline status to reflect that he was taller than definitely short people and shorter than people who were definitely non-short. Are we to suppose that  $\emptyset$  exists less than clearly existing items but more than clearly non-existent items? That is not an easy thing to suppose. On the one hand, there *are* no clearly non-existent items; to say that there is an item of a certain sort is to say that an item of that sort exists, and no non-existent items exist.<sup>5</sup> And on the other hand, if  $\emptyset$  is available for this sort of comparison at all, then it just simply exists, no maybe about it. You are welcome to look for another model, but I submit that it is incomprehensible how  $\emptyset$  or anything else could be without a determinate ontological status. One can't make sense of the notion that an existence question should be objectively unanswerable—unanswerable not just in the sense that we can't know the answer, but in the sense that there is no answer to know.

Q. The challenge is to explain why ontological mootness is not irremediably incoherent. I suppose I will have to do as you suggest and look for another

' $\emptyset$ ' is a definite mathematical symbol; it's a *circle* with a slash through it, not a '0' with a slash through it!

<sup>4</sup> Not always; but enough for present rhetorical purposes.

<sup>5</sup> See (Azzouni 2004) and (Hofweber 2005, 256–83) for the view that quantification has less to do with existence than is commonly thought.

model—an alternative to the borderline case model. Of course, it is one thing to show how ontological mootness could work in principle, another to show how it works in actual fact. I take you to be charging me with the former task. A model will eventually be suggested, but I will not be arguing that it is correct, only that it gives a way of making sense of the ontological mootness hypothesis. It is meant to illustrate the *kind* of work that needs to be done if we are ever to graduate from the wide-eyed wonderment stage of our ontological development.

*O.* That is a modest goal indeed. It won't bother you if the model is completely far-fetched?

*Q.* I said I won't be arguing that the model is correct. But it shouldn't be obviously incorrect either. There are two opposite dangers to be avoided: casting our net too wide, and not casting it widely enough. The model should not pull the rug out from under *all* existence-questions, because many such questions would seem to make perfect sense.

Desideratum 1: Where there is plausibly a fact of the matter about the existence of entities of type *Z*, the model should leave that fact in place.

Insofar as the answer to “Does the planet Vulcan exist?” is plausibly NO, and the answer to “Does the planet Earth exist?” is plausibly YES, we should take care that our model does not apply to planets, nor (I would tentatively suggest) to other commonsense macro-objects. Neither, though, should the model clearly *not* apply to existence-questions of a type apt to arouse quizzicalist suspicions.

Desideratum 2: If there is a type of entity *Y* whose ontological status is apt to seem moot, there should be the potential at least for bringing the model to bear on *Y*s.

Insofar as the ontological status of sets, say, or possible worlds, or random mereological sums, is apt to appear moot, our model should have the potential to vindicate that appearance. Putting these goals together, we want to show how there could fail to be a fact of the matter about whether *X*s exist, where the explanation *doesn't* generalize to questions about *Z*s, and *might or might not* generalize to questions about *Y*s.

### 3. ONTOLOGY RECAPITULATES PHILOLOGY

*O.* Let's first remind ourselves how the debate goes that you are alleging is empty. I have a sense your doubts reflect an overambitious idea of what it means for a thing to exist. Listen for a bit to my friend the platonist.

*P.* Ontology has been evolving, *Q*; we platonists, at least, are not as naïve as you think. Time was when ontological issues, about the existence of, say, *numbers*, were considered prior to linguistic issues, about the semantics of terms

purporting to *refer* to numbers. This led to ridiculous skeptical worries about numbers' existence that no amount of referential-seeming behavior on the terms' part could fully allay. Those days are gone. We have no use for

the possibility of some sort of independent, language-unblinkered inspection of the contents of the world, of which the outcome might be to reveal that there was indeed nothing there capable of serving as the referents of . . . numerical singular terms (Wright 1983: 13–14).

Contributing in a systematic, distinctive way to the truth-values of its containing sentences is *all there is* to a term's referring. Numerical and set-theoretic terms do this, so they refer, so numbers and sets exist. End of story.

*N.* I couldn't help overhearing. "The terms contribute, so they refer, so the objects exist: end of story." That will come as news to Scotland Yard. The view there is that Sherlock Holmes doesn't in fact exist, the truth-value-affecting powers of "his" name notwithstanding.<sup>6</sup> Astronomers will be interested to learn that there is a tenth planet, due to the distinctive effects of "Vulcan" in contexts like ". . . is supposed to be closer to the sun than Mercury".

*P.* I grant that we experience statements "about Holmes" and "about Vulcan" as true/false. But there might be various reasons for this. If indeed the terms are non-referring, then the statements are true only in a certain story or pretense or myth—not in real life. If you insist they *are* true in real life, then I reply that Holmes and Vulcan must exist after all, not of course with the properties they are represented as having in the story (pretense, myth), but as special sorts of abstract artefact: "fictional character," "failed posit."<sup>7</sup> The point is that you are really walking a fine line here. What you would need to make the objection work is a term that, one, has nothing make-believy about it; two, does not refer in anyone's book; and three, nevertheless makes a distinctive semantic contribution.

*N.* Well, I seem to remember that there is one expression whose whole philosophical *raison d'être* has been to serve as an example of a "serious" term with no real-world correlate. I mean, of course, "the King of France."<sup>8</sup>

*P.* You're joking, right? "The KoF" makes a distinctive semantic contribution? Not according to Russell or Strawson, it doesn't. The one maintains that sentences containing an empty description are one and all false.<sup>9</sup> The other thinks the question of truth or falsity does not even arise for such sentences. They agree that substituting one empty description for another leaves truth-status unchanged, as does changing the predicative context in which the description occurs. They

<sup>6</sup> (Divers and Miller 1995, 127–39).

<sup>7</sup> (Kripke's 1973 Locke Lectures) (Salmon 1998, 277–319) (Thomasson 1999).

<sup>8</sup> This is to construe "term" broadly, as befits the Fregean platonist context of the discussion.

<sup>9</sup> Strictly speaking, this applies only to empty descriptions in primary position; I will treat the restriction as tacit.

agree that the truth-status of an empty-description sentence (where, in Russell's case, the description gets wide scope) is fixed by the fact that it contains an empty description.

*N.* Yes, well, they may agree, but are they right? Strawson came to have doubts about this, because of examples like the following:

Suppose . . . that I am trying to sell something and say to a prospective purchaser, *The lodger next door has offered me twice that sum*, when there is no lodger next door and I know this. It would seem perfectly correct for the prospective purchaser to reply *That's false*, and to give as his reason that there was no lodger next door. And it would indeed be a lame defense for me to say, *Well, it's not actually false, because, you see, since there's no such person, the question of truth and falsity doesn't arise* (Strawson 1954: 225).

To the extent that "The lodger next door offered me twice that sum" strikes us as false, empty-description sentences are not all alike.

*P.* Come on. "That's false" is just a way of calling the landlord deceitful. Give me an example without the moralistic distractions. Give me an example with "the King of France."

*N.* I will try. Strawson is right that we are reluctant to take a stand on

- (1) The KoF is bald.

But we feel no such hesitation with

- (2) The KoF is sitting in that chair (pointing).

(2) is, at the very least, *a great deal less satisfactory* than (1). We are also not much put off by

- (3) The KoF has never worn these pajamas.

(3) is a great deal *more* satisfactory than "The KoF is bald." I would go so far as to say that (2) strikes us immediately as *something very like false*, and (3) strikes us as *something very like true*. Perhaps we don't want to insist that (1)–(3) differ in truth-value properly so called; Strawson certainly didn't. Still, we need to mark these differences somehow; so let us say that (2) *counts* as false, and (3) *counts* as true.

*P.* Hmmm. That's interesting as a parlor trick. But it doesn't change my view about numerical terms. I think I will just concede to you that a term like "the King of France" can in limited ways affect a sentence's felt truth-value. This doesn't bother me since any influence here is quirky and unsystematic—as we can see from the fact (noted by Strawson) that although (1) *normally* strikes us as unevaluable, as a response to "What bald notables are there?" it seems false. "The KoF" may be slightly semantically influential, but its influence falls far short of what we get with numerical terms like "the number of planets."

## 4. COUNTING AS FALSE

*N.* How quirky and unsystematic the contribution is remains to be seen. It is just not obvious at this point what is involved in a sentence's counting as false, or true, or neither. I say we try to figure it out. There are two ways to conceive the task, depending on whether we think "The A is B" is undefined in the absence of an A, or false. It will be simpler to take a Strawsonian line. The question then becomes, Why would a sentence  $\varphi$  that is undefined due to presupposition failure nevertheless strike us as false? An answer suggests itself almost immediately.  $\varphi$  might entail *another* sentence  $\psi$  whose (weaker) presuppositions are satisfied, allowing it to be genuinely true or false; and this other sentence  $\psi$  might be genuinely false. Small wonder if a sentence entailing a falsehood strikes us as false itself. "The KoF is sitting in this chair" counts as false because it entails that *someone* is sitting in the chair, when we can see that the chair is empty.

*P.* Hold on. If "The KoF is sitting in this chair" counts as false by virtue of entailing that the chair is occupied, shouldn't "The KoF is bald" *also* count as false, by virtue of entailing that France has a king?

*N.* Ah, but "The KoF is bald" does *not* entail that France has a king, when entailment is properly understood. It can't be regular old implication for just the reason you give: every KoF-sentence would count as false by virtue of having a false presupposition. That was not the idea! To entail a falsehood,  $\varphi$  must imply a false  $\psi$  whose falsity does not merely reflect the fact that  $\varphi$ 's presupposition  $\pi$  is false.  $\varphi$  counts as false just when there is a falsehood among its  $\pi$ -free implications. "The KoF is bald" does imply that France has a king, but that France has a king is not  $\pi$ -free; it is not free of the presupposition that France has a unique king.

*P.* Free in what sense, exactly?

*N.*  $\psi$  is  $\pi$ -free if it is false for reasons independent of  $\pi$ —for reasons that could still have obtained even had  $\pi$  been true. If by a falsity-maker for  $\psi$  we mean a truth-maker for its negation,  $\psi$  is  $\pi$ -free iff its falsity-makers are always compatible with  $\pi$ .<sup>10</sup>

<sup>10</sup> What is it for something to be the reason why  $\varphi$  is true in a world? Truth-makers for our purposes can be true propositions—so in one sense of the word, 'facts'—implying the proposition that  $\varphi$ , and thus guaranteeing that  $\varphi$  is true. But which  $\varphi$ -implying fact is (facts are) to play the role of its truth-maker(s)? I take the word "maker" here to indicate that we are interested in a fact that in some sense *brings it about* that S is true. And so I look for inspiration to another kind of bringing about, viz. causation. There are two desiderata that have to be traded off against one another when we try to pick an event's cause out from among the various antecedents that are in the circumstances sufficient for it.

One is *proportionality*: the cause shouldn't contain too many extra details in whose absence you'd still have an event sufficient for the effect. Getting hit by a bus is a better candidate for cause of death than getting hit by a red bus (an example of Williamson's). Proportionality cannot be pursued at all costs though, or we will wind up with disjunctive causes whose disjuncts correspond to all

*P.* Examples, please—preferably not involving the King of France this time.

*N.* Right. Remember, the claim is that  $\varphi$  counts as false iff it has false implications, where the falsity is for  $\pi$ -compatible reasons. Consider

(4) Both of Bush's wives are Jewish.

This presupposes that Bush has two wives, but that can't be the reason it strikes us as false. It strikes us as false because it implies, for example, that all of Bush's wives are Jewish, and this is false for a  $\pi$ -compatible reason, viz. that Laura is a Bush wife and Laura is not Jewish. The implication is  $\pi$ -free because for Laura to be a non-Jewish Bush wife is fully compatible with Bush having another wife off to the side, making  $\pi$  true. "Both of Bush's wives are Jewish" counts as false by virtue of  $\pi$ -free implying that among Bush's wives one finds only Jews.<sup>11</sup>

## 5. COUNTING AS TRUE

*P.* OK, but more interesting is the case where a sentence whose presuppositions fail nevertheless strikes us as *true*. Let me guess:  $\varphi$  counts as true iff its negation counts as false.

*N.* Not quite. The problem is that nothing so far rules out a sentence and its negation *both* counting as false; and if any such cases occur, the proposed account will lead us to count  $\varphi$  true and false at the same time. Such cases do seem to occur.

possible ways for the effect to come about: getting hit by a bus or a car or a plane or having a piano or safe fall on your head or etc. The other desideratum, then, is *naturalness*: given a choice between two otherwise qualified candidates for the role of cause, we prefer the less disjunctive one. These desiderata pull against each other because the less disjunctive a prior event becomes, the more it contains "extra" details, in whose absence it would still have been sufficient. But some tradeoffs are better than others, and what we look for in a cause is a prior event that effects the best tradeoff possible: you cannot make it more proportional except at the cost of a whole lot of disjunctiveness, and you cannot make it less disjunctive without making it a whole lot less proportional.

With truth-makers, too, the first desideratum is *proportionality*: one wants a proposition that doesn't contain extra details in whose absence you'd still have a proposition implying that  $\varphi$ . So, "Someone is sleeping" is not made true by the fact that Zina is sleeping *fitfully*; the fitfulness is irrelevant, the fact that Zina is sleeping is enough. The second desideratum is *naturalness*: one wants a proposition that is not unnecessarily disjunctive. "Someone is sleeping" is made true by the fact that Zina is sleeping, not the fact that Zina is sleeping or Vanessa is sleeping or Ruth is sleeping or etc. A truth-maker for sentence  $\varphi$  is a  $\varphi$ -implier that effects a better tradeoff between proportionality and naturalness than relevant competitors.

<sup>11</sup> "Shouldn't 'The king of France is bald' count as false, by virtue of implying (not 'France has a king' this time, but) 'France has a bald king?'" The answer is that 'France has a bald king' is not  $\pi$ -free; it is made false by the  $\pi$ -incompatible fact that France has no king. "Who's to say the falsity-maker isn't France's lack of a *bald* king? France's lacking a bald king is fully  $\pi$ -compatible." France's lack of a bald king does not effect the best tradeoff between proportionality and naturalness. The baldness is an unnecessary complication; one can strike it and still be left with a fact—France's lack of a king—implying the falsity of 'France has a bald king.' France's lack of a king is also the more natural (because less disjunctive) of the two; there are more ways for France to be without a bald king than without a king. The falsity of 'France has a bald king' is better blamed on France's lack of a king than its lack of a bald king.

(5) The author of *Principia Mathematica* was bald

counts as false by virtue of implying the  $\pi$ -free falsehood “All *PM* authors were bald.” (One of *Principia Mathematica*’s two authors was Bertrand Russell, who had a full head of hair late into his life.) (5)’s Strawsonian negation

( $\sim$ 5) The author of *Principia Mathematica* was non-bald

counts as false for a similar reason. It  $\pi$ -free implies that all *PM* authors were non-bald, which is refuted by Alfred North Whitehead’s (white) head. If ( $\sim$ 5)’s counting as false conferred on (5) the property of counting as true, then (5) would count as true and as false at the same time. Clearly, though, that is not how (5) strikes us. This suggests that for  $\varphi$  to count as true it is required that  $\sim\varphi$  counts as false *while  $\varphi$  itself does not count as false*.  $\varphi$  counts as true iff its negation implies  $\pi$ -free falsehoods, *while  $\varphi$ ’s  $\pi$ -free implications are one and all true*.

*P.* And  $\varphi$  counts as gappy iff it counts neither as true nor as false?

*N.* That’s right. Take (1) = “The KoF is bald.” (1) doesn’t count as false because although it has false implications, e.g., *some French king is bald* they are not  $\pi$ -free; they are false for reasons that require  $\pi$  too to be false. Its  $\pi$ -free implications, for instance, *all French kings are bald* are all true. The argument that its negation ( $\sim$ 1) doesn’t count as false is similar. ( $\sim$ 1)’s implication that *some French king fails to be bald* is false, but not  $\pi$ -free, while its implication that *all French kings fail to be bald* is  $\pi$ -free but not false. That ( $\sim$ 1) does not count as false means that (1)—already seen not to count as false—does not count as true either. So it counts as gappy.

*N* (continuing).  $\varphi$  counts as gappy iff its  $\pi$ -free implications are all true and the same holds of its negation. This makes good intuitive sense; a sentence so tainted by its association with  $\pi$  that there is nothing left for it and its negation to disagree about when  $\pi$  is stripped away is not making an evaluable claim. When  $\varphi$  makes no claim as a result of  $\pi$ ’s failure, we say its presupposition *fails catastrophically*. This is the case where, as Strawson puts it, “the whole assertive enterprise is wrecked” by  $\pi$ ’s falsity. It is the other, non-catastrophic, sort of presupposition failure that I am concerned to emphasize today.

## 6. NOMINALISTIC RAMIFICATIONS(?)

*P.* Emphasize away. I am still not clear what the bearing of non-catastrophic presupposition failure is supposed to be on the issues that divide us.

*N.* Haven’t we already discussed this? *You* said you believed in numbers because of the truth-value-affecting powers of numerical terms. *I* said that the power of affecting truth-value is not reserved to referring terms, witness “Holmes” and “Vulcan.” *You* said that to the extent these terms affect truth-value, they refer; to the extent that they do not refer, they affect only truth-value-in-the-story. *You*

demanded an example of a semantically influential term with nothing make-believly about it, and which *clearly* fails to refer. *I* gave you one: “the King of France.”

*N* (continuing). The look on your face tells me that you are still confused. Let me spell it out for you. If the untruth of a *concrete* existence presupposition is compatible with a sentence’s properly striking us as true or false, why not also the untruth of an *abstract* existence presupposition? I suggest that

(6) The number of planets is odd

strikes us as true for the same sort of reason as “The King of France sits in this chair” strikes us as false.<sup>12</sup>

*P*. You seem to have forgotten the ground rules. I asked for an example of a non-referring term that influences *truth*-value. Sentences containing “the King of France” may *count* as true or false, but that is not the same as *being* true or false.

*N*. It is *almost* the same, though. Here is why. Our notion of  $\pi$ -free implication bears affinities to a distinction that linguists make between what is *presupposed* in an utterance and what is *asserted* or *alleged* or *at issue*. If I say that both of my children play soccer, this presupposes that I have two children, and asserts that my children, whatever their number, play soccer. Let me now propose a

BRIDGE PRINCIPLE:  $\varphi$ ’s assertive (allegational, at-issue) content is the sum total of its  $\pi$ -free implications; equivalently  $\psi$  is part of  $\varphi$ ’s assertive (. . .) content iff  $\psi$  is a  $\pi$ -free implication of  $\varphi$ .<sup>13</sup>

The reason having a false  $\pi$ -free implication makes  $\varphi$  count as false is that something false is *asserted*. Recall that  $\varphi$  makes a claim iff at least one of  $\varphi$ ,  $\sim\varphi$  counts as false; let’s now add that *the claim  $\varphi$  makes* when this condition is met is its assertive content. Then the above explanations of counting as true (false, gappy) boil down to this:

(C)  $\varphi$  counts as true (false) iff  $\varphi$  makes a true (false) claim.  $\varphi$  counts as gappy iff it makes no claim.

Now let me reply to your charge that counting as true (false) is one thing, being true (false) is another. The charge is correct but it ignores that  $\varphi$ ’s *counting* as true (false) goes with the *genuine* truth (falsity) of the claim  $\varphi$  makes. That a KoF-sentence’s counting as true (false) is not the same as its being true (false) doesn’t matter, for it *is* the same as the truth (falsity) of the expressed claim. Empty terms affect the *genuine* truth-value of what is claimed, and that is enough.

<sup>12</sup> Assuming for sentimental reasons that Pluto is still a planet.

<sup>13</sup> The principle would need to be complicated to deal with cases where presupposed content is repeated in an assertive mode, as it is sometimes held that knowledge attributions presuppose and assert their complements. I will ignore this complication here, though it is important for the understanding of existence claims. “That little green man exists” arguably presupposes the man’s existence in the course of asserting his existence.

*N* (continuing). Where does this leave us? Sentences like (2) and (3) show that existential presupposition failure is no bar to a sentence's striking us as true or false—indeed to its *properly* striking us as true or false, since it seems only proper to evaluate a sentence on the basis of what it asserts, as against what it assumes as background. But then the failure of the existential presupposition in

(6) The number of planets is odd

is no bar to (6)'s properly striking us as true—indeed to its *being* true in the sense that what it asserts is true. To repeat my suggestion above, NoP-sentences (“number of planet”-sentences) are true (false) in the same way as, and to the same extent that, KoF-sentences like (2) and (3) are true (false).

*P*: So you say, but you have to admit that the cases look disanalogous. I grant you that

(i) *positive* empty-description sentences like “The KoF is sitting in this chair” are apt to strike us as *false*.

I grant you that

(ii) *negative* empty-description sentences like “The KoF will never wear these pajamas” are apt to strike us as *true*.

I will even grant you that

(iii) *positive overfull*-description sentences like “The author of *Principia Mathematica* was good at math” are apt to strike us as *true*.

But we have not seen any examples of

(iv) *positive empty*-description sentences (of the same form as (6) “The number of planets is odd”) that strike us as *true*.<sup>14</sup>

And that is what we will need if we are to use the King of France (KoF) as a model for the number of planets (NoP). For it hardly needs saying that positive sentences about the number of planets quite *often* strike us as true.

*P* (continuing). Before I'll accept that positive number-presupposing sentences like (6) can count as true in the absence of numbers, you will have to give me an example of a positive French-king-presupposing sentence that counts as true in the absence of French kings. On the face of it, there do not seem to be any. “The KoF is bald” counts as gappy, and “The KoF is sitting in this chair” counts as false; where are the positive predicates such that “The KoF is P” counts as true?

*N*. I think you are misunderstanding my argument—or maybe I misrepresented it. You seem to think I am offering an argument by analogy: numerical terms

<sup>14</sup> The few examples that come to mind are conceptual truths about French kings as such, or else intentional ascriptions with “the KoF” appearing in an opaque position.

make the same sort of contribution as “the King of France”; the latter doesn’t refer; so the former don’t (or needn’t) refer either.

*N* (continuing). But that is not the argument at all. “The King of France” was brought in not as a model for “the number of planets,” but to motivate a certain theory of non-catastrophic presupposition failure: the theory stated above. The argument that (6) still counts as true in the absence of numbers is not that that’s what you’d expect based on the analogy with (2) and (3), it’s that the theory assigns

- (7) There is exactly one planet or there are exactly three planets or there are exactly five planets or etc. . . .

to (6) as its assertive content;<sup>15</sup> and since (7) clearly depends for its truth just on how many planets there are, whether (6) counts as true depends just on how many planets there are.

## 7. PLATONISTIC RAMIFICATIONS(?)

*P*. Say you’re right that (6) still counts for us as true whether numbers exist or not. That doesn’t show numbers do *not* exist. The most it shows is that (6)’s counting as true *leaves it open* whether numbers exist.

*P* (continuing). The question we should be asking is: Is there anything special about the way numerical terms influence felt truth-value to suggest either that they do refer, or that they don’t? It seems to me that the abundance of positive truths like (6)<sup>16</sup> now becomes relevant again. If a concrete term’s emptiness manifests itself in a shortage of positive truths, why wouldn’t an abstract term’s emptiness manifest itself the same way? Numerical terms figure in plenty of positive truths, however. Related to this, a concrete term’s emptiness manifests itself in an abundance of truth-value gaps; Strawson’s theory would never have got off the ground if KoF-sentences did not frequently strike us as unevaluable. NoP-sentences hardly ever strike us this way. Numerical terms exert a much stronger semantic influence than the empty terms we know best: empty concrete terms. Come to think of it, they are about as semantically influential as referring concrete terms. These observations about pattern of influence suggest that numerical terms refer.

*N*. I grant you that empty concrete terms figure in few intuitive truths, and many intuitive gaps. The contrast with numerical terms could not be more striking.

<sup>15</sup> (6) clearly implies (7). The implication is  $\pi$ -free because a falsity-maker for (7) would be a fact to the effect that there are so and so <sup>many</sup> planets, and facts about planets cannot conflict with (6)’s presupposition that there are numbers. So (7) is part at least of (6)’s assertive content. I can’t rule out that (6) has other  $\pi$ -free implications, making for a stronger assertive content, but if so I don’t know what they would be.

<sup>16</sup> I am using “truths” now not for true sentences but sentences making true claims. Likewise “falsehoods.” “Gaps” are sentences making no claim.

But there are two possible explanations of this contrast: one is that numerical terms are not empty; another is that *the emptiness of a numerical term is much less of a drain on its semantic influence than the emptiness of a concrete term.*

*N* (continuing). The second explanation seems more plausible. A King of France, if there was one, would be an original source of information of the type that makes KoF-sentences true; our presumptions about what a French king would have to be like are far too weak to take up the slack in his absence. Numbers if they existed would *not* be an original source of information of the type to make numerical sentences true. Our presumptions about how numbers, if they existed, would relate to other things *are* in this case enough to take up the slack. Indeed, that ~~would seem to~~ <sup>might</sup> be what distinguishes concrete terms from abstract: it holds of concrete terms but not abstract that whether the term refers makes a large difference to which of its containing sentences count as true and false.

*P*. I might even concede to you that numerical terms' effect on the distribution of truth-values doesn't *require* them to refer. But you haven't convinced me that they don't refer anyway, non-obligatorily as it were. It may not be the only way to make sense of their semantic impact, but it's certainly the most natural.

*N*. And I concede to you that numerical terms' effect on the distribution of truth-values doesn't pattern with that of empty concrete terms. Still, you haven't convinced me that this is evidence of non-emptiness as opposed to abstractness. The semantic powers of an abstract term are exhausted by what the referent is *supposed* to be like, and that remains in place whether the referent is there or not.

## 8. QUIZZICALISTIC RAMIFICATIONS(?)

*Q*. I find these concessions suggestive. Let me restate them in different terms. A mechanism is fail-safe, according to my dictionary, if the surrounding system is "capable of compensating automatically and safely for its failure." Existential presupposition is a mechanism in the larger machinery of assertion and fact stating—a mechanism that is liable to fail. The failure of *concrete* existential presuppositions is often (although, we have seen, not invariably) catastrophic; if I describe the F as G, and there turn out not to be any Fs, then the assertive enterprise is often, as Strawson says, "wrecked." The failure of *abstract* existential presuppositions is generally *non-catastrophic*, however. (6) still makes a claim about how many planets there are even if there are not any numbers—a claim that is part and perhaps all of the one that would have been made had numbers existed.<sup>17</sup> *Abstract presuppositions are, to that extent, a fail-safe mechanism within the larger machinery of assertion.* It is not that the mechanism can't fail,

<sup>17</sup> "Part" if we suppose, as I generally do not in this paper, that we scale  $\varphi$ 's full content back to its  $\pi$ -free content only on condition that  $\pi$  is false.

but that the failures don't *matter*, as the machinery of assertion "compensates automatically and safely" when they occur.

Q (continuing): Now, suppose the platonist is right that whether a term (an abstract term, anyway) refers is entirely a function of the term's sentence-level semantic effects—its effects on what is claimed and on whether the sentence counts as true, false, or gappy.<sup>18</sup> And suppose the presupposition that it *does* refer is fail-safe: the term's sentence-level semantic effects are the same whether it refers or not. If the one factor that is available to determine whether numerical terms refer takes the same value whether they refer or not, then that factor is powerless to settle whether numerical terms refer. By hypothesis, though, semantic influence *is* the only determining factor; if it fails to settle whether numerical terms refer, then nothing settles it, and the matter is objectively unsettled.

O. You keep on talking about numerical *terms*. But the issue between us isn't a metalinguistic one; it concerns the numbers themselves. Let it be that there is no fact of the matter as to whether numerical terms refer. There could still be a fact of the matter as to the existence of numbers.

Q. Could there really? In practice, the issues are very hard to tell apart. Ask yourself what it would be like to think that although 2 determinately existed, it was indeterminate whether "2" referred. One would have to think that "2" did not determinately refer to 2. (How could it be indeterminate whether "2" referred, if there determinately existed a thing that was determinately its referent?) This idea of 2 determinately existing but eluding semantic capture by "2" is hard to make sense of. Again, what would it be like to think that although 2 determinately failed to exist, it was indeterminate whether "2" referred? One would have to think that there was something other than 2 such that "2" did not determinately fail to refer to this other thing. (How could it be indeterminate whether "2" referred, if its one candidate referent determinately failed to exist?)

Q (continuing). The issue of whether numbers exist is hardly to be distinguished from the issue of whether numerical terms refer; if the one is objectively unsettled, as we have said, then the other is objectively unsettled too. This is how there can fail to be a fact of the matter about numbers' existence.

O. You said that you were going to offer a *model*. So far, all I am seeing is an example.

Q. Suppose that our interest in *Xs* stems mainly from the role *X*-expressions play in sentences of a certain type: *X*-sentences, let's call them.<sup>19</sup> Suppose that

<sup>18</sup> Terminology is confusing here, because for a term to refer to so and so does not imply that it refers simpliciter. To say that a term *t* refers is to make an existential claim to the effect that it has a referent. To say that *t* refers to so and so is to say that its referent is so and so on the assumption that it refers.

<sup>19</sup> I assume that the line taken above about numerical definite descriptions can be extended to numerals, numerical quantifiers, and the like.

$X$ s are presupposed by  $X$ -sentences and that the presupposition is fail-safe in the following sense: if  $\varphi$  is an  $X$ -sentence, then  $\varphi$ 's assertive content is the same, and has the same truth value, whether  $X$ s exist or not. Then there is nothing to determine whether the  $X$ -expressions in  $X$ -sentences refer, and to that extent, nothing to determine whether  $X$ s exist.

*O.* That tells me when (in your view) ontological mootness arises, but not how it is possible in the first place. How does your model address the *mystery* I was complaining about it in the first section?

*Q.* Well, how does the vagueness of “short” remove the mystery of there being no fact of the matter as to whether Tom Cruise is short? I take it the explanation goes something like this. Vagueness is semantic underspecification. Gather together all the factors that are supposed to determine the extension of “short”; you will find that they constrain the extension, but do not succeed in determining it completely. The extension of “short” is to that extent unsettled; and for the extension to be unsettled is no different from its being unsettled who is short.

*Q* (continuing): The explanation of how it can be unsettled whether there are  $X$ s is broadly similar, especially in the use it makes of semantic underspecification. The factors that are supposed to determine which expressions refer do not always succeed in doing this. They constrain which expressions refer but they do not determine it completely; they do not determine whether expressions refer which have the same sentence-level effects regardless. Since there is by hypothesis nothing else to determine whether these expressions refer, there is no fact of the matter either way. For there to be no fact of the matter whether  $X$ -expressions refer is the same as there being no fact of the matter whether  $X$ s exist.

*O.* There is still the mystery of why a language should contain expressions whose referential status it is content to leave undetermined.

*Q.* Let me try a just-so story out on you. I start from the fact that presuppositions are not on the whole advanced as true. It makes sense, then, that the mechanisms driving semantic evaluation would try their best to bleach presuppositional content out and focus on  $\pi$ -free implications, or what I have called assertive content; assertive contents should ideally evaluate the same whether  $\pi$  is true or not. Most terms, however, and certainly most concrete terms, will not submit to this treatment; as we saw with “the King of France,” they by and large enable the expression of interesting (in particular, true) claims only if they refer. BUT: *terms could evolve that play into the presupposition-discounting mechanism*, engendering the same claims, with the same truth-values, regardless. I suggest that numerical terms, and abstract terms more generally, are like this. They evolved to influence what is said by virtue of non-referential properties only—by virtue of the kind of thing they are *supposed* to pick out. Ontological mootness is a natural if unintended by-product.

## 9. EXTENT OF THE PHENOMENON

O. It sounds like you are saying that it is when  $X$ s are or would be abstract that there is no fact of the matter about their existence.

Q. Not really. There is no fact of the matter about the existence of  $X$ s when the presupposition that they exist is *fail-safe* in the sense discussed. I agree, though, that it is generally with abstract objects—numbers, sets, truth-values, shapes, sizes, amounts, and so on—that this happens.

O. That is excellent news. It means that as long as I restrict my attention to *concrete* objects, you will have no ground for complaint. That should still leave me with plenty to do; I can worry myself about Lewisian possible worlds, the equator, the twentieth century, and the mereological sum of ~~my~~ one pair of dress pants with its matching jacket. <sup>your</sup>  $\wedge$

Q. I might still complain. Take “The mereological sum of my pants and jacket is at the cleaner’s.” Stripped of the presupposition that my pants and jacket have a mereological sum, this says that my pants and jacket are at the cleaners.<sup>20</sup> The assertive content is the same whether the mereological sum exists or not, and its truth-value is the same, too. If this pattern continues, then the existential presupposition is fail-safe, and there is no fact of the matter as to whether my pants and jacket have a mereological sum.

O. You said earlier that your model aspires to pull the rug out from under some ontological questions, but not all. It’s beginning to sound like you want it to pull the rug out from under “philosophical” existence questions, but not “ordinary” ones about commonsense objects like, I suppose, pants.

Q. Go on.

O. Well, it seems to me the model applies just as well to pants as to sums of pants and jackets. Take “<sup>You</sup> I have a pair of pants at the cleaner’s,” or to avoid the complexities of ownership, “There is a pair of pants at the cleaner’s.” Stripped of the presupposition that the microparticles in pants have a mereological sum, this says that pantishly arranged microparticles are at the cleaner’s. The assertive content is the same whether the pants exist as a further entity or not, and its truth-value is the same too. Apparently there is (going by your criterion) no fact of the matter as to whether pants exist either.

<sup>20</sup> I take it that “The mereological sum of  $x$  and  $y$  is at the cleaner’s” ( $\varphi$ ) implies “ $x$  and  $y$  are at the cleaner’s” ( $\psi$ ) and presupposes “ $x$  and  $y$  have a mereological sum  $x+y$ ” ( $\pi$ ).  $\psi$  is  $\pi$ -free to the extent that falsity-makers for “ $x$  and  $y$  are at the cleaner’s” speak only to  $x$  and  $y$ ’s locations and do not conflict with  $x$  and  $y$  having a mereological sum. So  $\psi$  is at least part of  $\varphi$ ’s assertive content. I can’t rule out  $\varphi$ ’s having additional  $\pi$ -free implications, making for a stronger assertive content, but I don’t know what they would be.

Q. You say “There is a pair of pants at the cleaner’s” has “Pantishly arranged microparticles are at the cleaner’s” as its assertive content. That is just not so on my view. Remember,  $\varphi$ ’s assertive content is made up of its  $\pi$ -free *implications*. Implications (I could have made this clearer) are statements whose truth follows *analytically* from the truth of their impliers. The microparticle-statement figures in the assertive content of the pants-statement only if it is analytically implied by the pants-statement. But the pants-statement does not analytically imply there *are* such things as microparticles, let alone that there are pantishly arranged microparticles at the cleaner’s. I deny, then, that “There is a pair of pants at the cleaner’s” has an assertive content that concerns just microparticles. There being no fact of the matter about mereological sums does not preclude a fact of the matter about regular macro-objects.

O. I concede that the pants-statement does not imply anything about microparticles. But it does have *some* sub-pant implications; it implies, for instance, that there are pant-legs at the cleaners.<sup>21</sup> Perhaps “There are pants at the cleaner’s” has an assertive content to do with pant-legs! Won’t that assertive content be the same, and retain the same truth-value, whether pants exist or not?

Q. Here I have to remind you of a crucial distinction.  $\varphi$ ’s assertive content is the sum  $\alpha(\varphi)$  of its  $\pi$ -free implications.  $\alpha(\varphi)$  does not count as “the claim  $\varphi$  makes,” though, unless it conflicts with  $\alpha(\sim\varphi)$ ; at least one is false. The problem with (1) = “The King of France is bald” is not that it lacks an assertive content; it’s that no claim is made since  $\alpha(1)$  and  $\alpha(\sim 1)$  are both true. For  $\pi$  to be fail-safe, it’s not enough that  $X$ -statements  $\varphi$  have the same assertive contents, with the same truth-values, whether  $\pi$  holds or not; they have to make the same *claims*, with the same truth-values.

O. How does that bear on the issue of pants?

Q. Let  $\varphi$  = “There are pants at the cleaner’s.”  $\alpha(\varphi)$  is the sum total of what “There are pants at the cleaner’s” implies about how matters stand pants aside.  $\alpha(\varphi)$  counts as a claim  $\varphi$  makes only if it conflicts with  $\alpha(\sim\varphi)$  = the sum total of what “There are *not* pants at the cleaner’s” implies about how matters stand pants aside.  $\varphi$  makes a pants-free claim, in other words, only if “There are pants at the cleaner’s” and “There are not pants at the cleaner’s” have conflicting analytic implications for what goes on at lower levels of reality. What would they be? I am willing to grant that “There are pants at the cleaner’s” analytically implies that there are pant-legs at the cleaner’s. But this doesn’t get us a  $\pi$ -free *claim* unless “There are *not* pants at the cleaner’s” analytically implies that there are *not* pant-legs at the cleaner’s. And there is clearly no such implication. One way for there to be no pants at the cleaners is for there to be no pant-legs there. Another way is for the pant-legs there to be unmatched and unattached.

<sup>21</sup> Thanks here to Eli Hirsch.

Q (continuing). You are of course welcome to argue that there is more to  $\alpha(\varphi)$  and  $\alpha(\sim\varphi)$  than I have acknowledged, and that  $\varphi$  and  $\sim\varphi$  really do have conflicting analytic implications for the sub-pants order of things. That indeed seems like a good thing to try to argue. Myself, I doubt conflicting pants-free implications can be found. But I've been wrong before.

O. I have a better idea now where the debate is going; let's talk more tomorrow. Until then, explain one last time how we tell on your view when an existence-question is moot. I know the formula: "Are there  $X$ s?" is moot iff the presupposition of  $X$ s is fail-safe. But tell me again how the formula is to be understood.

Q. "Are there  $X$ s?" is moot iff hypotheses  $\varphi$  that presuppose  $X$ s<sup>22</sup> are systematically equivalent (modulo  $\pi$ ) to hypotheses  $\alpha(\varphi)$  about how matters stand  $X$ s aside.<sup>23</sup>

Q (continuing). You get the stated equivalence with sets, numbers, sizes, shapes, amounts, chances, possible worlds, and mereological sums. ("The amount of water in this pond exceeds the amount of water in that one" is equivalent modulo  $\pi$  to "There is more water in this pond than in that one." "There is a possible world where pigs fly" is equivalent modulo  $\pi$  to "It is possible for pigs to fly.") The model predicts, then, that it should strike us as moot whether sets, numbers, sizes, etc. really exist—or at least as mooter whether they exist than whether "regular" things like dogs and motorcars exist. "Regular" things are distinguished by the fact that statements about them are not systematically equivalent, modulo the assumption of their existence, to statements about anything else.

## REFERENCES

- Azzouni, J. 2004. *Deflating Existential Consequence: A Case for Nominalism*. Oxford: Oxford University Press.
- Burgess, J. and G. Rosen. 1997. *A Subject With No Object*. Oxford: Clarendon Press.
- Divers, J. and A. Miller. 1995. "Minimalism and the Unbearable Lightness of Being", *Philosophical Papers* 24 (2): 127–39.

<sup>22</sup> A bit more carefully: when we look at the hypotheses that lead us to take  $X$ s seriously in the first place, we find that (i) they presuppose the existence of  $X$ s rather than asserting it, and (ii) they are systematically equivalent (modulo  $\pi$ ) to hypotheses  $\alpha(\varphi)$  about how matters stand  $X$ s aside. If we agree with Frege that "it is applicability alone which elevates arithmetic from a game to the rank of a science" (*Grundgesetze*, vol. II, sec. 91, p. 187 in Geach and Black, 1960),  $\varphi$  would in the case of numbers be a hypothesis of applied arithmetic rather than pure. (I am *not* assuming that  $\alpha(\varphi)$  is straightforwardly expressible in English; we know it, in many cases, only as what  $\varphi$  adds to  $\pi$ .)

<sup>23</sup> "Systematically" in the sense that logical/conceptual relations are preserved.  $\varphi$  is inconsistent with  $\sim\varphi$ , for example, so  $\alpha(\varphi)$  should be inconsistent with  $\alpha(\sim\varphi)$ . This is the requirement Q is pressing in the main text, when he asks how  $\alpha(\text{There are pants at the cleaner's})$  conflicts with  $\alpha(\text{There are no pants at the cleaner's})$ .

9

- von Fintel, Kai. 2004. "Would You Believe It? The King of France Is Back! (Presuppositions and Truth-Value Intuitions)", in *Descriptions and Beyond*, ed. Reimer, M. and A. Bezuidenhout. Oxford: Clarendon Press.
- Frege, G., P. Geach, and M. Black. 1960. *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Basil Blackwell.
- Hofweber, T. 2005. "A Puzzle about Ontology", *Noûs* 39 (2): 256–83.
- Salmon, N. 1998. "Nonexistence", *Noûs* 32 (3): 277–319.
- Strawson, P. F. 1954. "A Reply to Mr. Sellars", *Philosophical Review* 63 (2): 216–31.
- Thomasson, A. L. 1999. *Fiction and Metaphysics*. Cambridge: Cambridge University Press.
- Wright, C. J. G. 1983. *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.
- Yablo, S. 2006. "Non-Catastrophic Presupposition Failure." In J. J. Thomson and A. Byrne (eds.) *Content and Modality: Themes from the Philosophy of Robert Stalnaker*. Oxford: Oxford University Press [Chapter 11 in this volume].

## Index

- abduction 248  
aboutness 146; *see also* unaboutness  
abstract objects 145–147, 148; *see also*  
    mathematical objects, arbitrary  
    mereological sums, models, numbers, sets  
a priori knowledge of 201–203  
equivalence relations and 258  
existence claims 122 n. 16  
figuralism and 208–9  
Goodmanian language and 241  
impure abstracta 205 n. 6  
necessary existence as  
    conservativeness 206–8  
necessary existence of 204–6  
presupposing existence of *see* abstract terms  
presupposition failure and 290–1, 310  
properties of 200–1  
quantification over 141 n. 73  
relation to nominalism and platonism 296  
    n. 3  
    as representational aids 195, 210–1  
abstract terms 291–93, 303–9  
abstraction principles 257; *see also* Hume’s  
    Principle  
accense 201  
accidental properties 14, 16 n. 6, 20–22,  
    25, 51, 200; *see also* hypothetical  
    properties  
    accidental extrinsic properties *see* extrinsic  
    properties  
    accidental intrinsics *see* intrinsic properties  
    and categorical properties 48, 65  
    and causally important properties 75, 77  
    and causation 74–5, 77  
    distinguished from contingent 71 n. 33  
    and essences 64, 65–6  
    and intrinsic properties 51  
    problem of accidental intrinsics; *see* intrinsic  
    properties  
accidental/essential distinction 65; *see also*  
    properties  
accidentalization 20–22  
accompaniment 33–5, 40, 53 n. 41  
actual scenario *see* branching time  
actual world 36, 53, 55, 56, 81, 82, 272  
    initial segment of actuality 39  
actuality *see* actual world  
adverbs  
    metaphorical potential of 161  
    and modification 68 n. 27, 155  
    semantics for 156  
aggregate 43  
    and compounding 48 n. 33  
    defined 45  
aggregates 43, 45–6, 48 n. 33, 57  
arbitrary mereological sums 296 n. 3, 298,  
    310–12  
arithmetic *see also* mathematics  
    absoluteness of 203, 219  
    applicability of 222–4  
    applied 182, 183, 191  
    arithmetical truth 194–5, 233, 292  
    existential commitments of 196  
    and fiction 182  
    necessity and a priority of 196, 201–203,  
    206–8, 233  
    pretense-worthiness of 233  
    real content of 211–13, 219  
    truth of 223, 230  
Arnauld, Antoine 16  
as-if 135–6, 163–4, 187, 188, 195, 196–7,  
    235  
    conception of sets 243  
    make as if to assert 179, 181, 183, 210, 211,  
    212, 214, 227, 230, 232, 238, 239,  
    241, 268; *see also* quasi-assertion  
    meaning 170–1  
attribute 20  
  
Basic Law V 257  
Bennett, Jonathan 64, 110  
brain state 86  
branching worlds 38 n. 13, 103–4  
    causes as choice-points 81  
    and overlap 39 n. 15  
bridge principles 146, 147–8, 149, 151, 156  
bridging a logical gap 187  
    bridge too far 286  
    and say-more strategy 283  
Burge, Tyler 258 **W.**  
Bush, George H. 235 **W.**  
Bush, Laura 302  
  
career  
    counterfactual 50, 71  
    modal 48  
    transworld 45, 205

- Carnap, Rudolf 119–30, 141 n. 73, 190  
 Cartesianism 93  
 carving content 246–7, 249; *see also*  
 abstraction principles  
 conflation of problem 255–6  
 inviting recarving 255, 260, 266–7  
 and numbers 267–8  
 proliferation problem 256–8  
 tolerating recarving 255, 260  
 and unwanted objects 260  
 categorical profile 50  
 categorical properties 24–31, 50–1, 59, 67  
 categorical duplicates *see* duplicates  
 and causal properties 70–1  
 and coincidence 25, 26, 29, 48, 65, 66–7  
 and contingent identity 23–4, 27, 31  
 and cumulative essence 29  
 and essential properties 68, 71  
 and functional properties 31  
 and intrinsicness 48–50, 52  
 and strengthening 66–7, 70  
 categorical/hypothetical distinction 31, 50, 65;  
*see also* properties  
 causal asymmetry 110–11  
 causal chain 100  
 causal consequence 94–5  
 causal essence 59, 68 n. 27, 71, 82  
 causally irrelevant properties in 72, 74,  
 75–6, 94  
 and proportionality 77  
 causal necessity 78, 80 n. 53, 85–6, 87–90, 94  
 causal ontology 77–81  
 causal power 32, 49–50, 64, 68, 71, 74, 77,  
 85, 91  
 causal properties 59, 61, 70–1, 76  
 and accidental properties 75, 77  
 competition between 91  
 causally irrelevant properties 72, 76  
 causal relevance 72, 84, 87, 90–92, 93  
 causal sufficiency 78, 84–5, 87, 90–93  
 causation *see also* cause  
 adequacy condition on 72–5, 76–7, 77–8  
 and accidents 80  
 causal judgment 82–3  
 causal origins 41  
 causal properties 23–4, 31–32, 48, 70–1  
 causal relations 81  
 contingency condition on 71–73, 75–6,  
 77, 78  
 counterfactual theories of 32, 72–3, 98,  
 99–100, 101, 106, 110, 113–14  
 and de facto dependence 105–6  
 and dependence 101–103  
 enoughness condition on 76–8, 81, 93–4,  
 98  
 and epiphenomenalism 83–90  
 Hume/Mill theory of 60  
 mental 31–32, 83–90, 92–4  
 physical 94–5; *see also* determinism  
 requirement condition on 75–8, 80–82, 93  
 n. 82, 94–5  
 singular and general 60 n. 7  
 and truth-makers 301 n. 10  
 types of causal power 50  
 causes  
 and accidents 80  
 as adequate for effect 72–5  
 as commensurate with effect 59–60  
 competition between causes 87, 91–92,  
 93–4  
 as concrete 62  
 constitutionalism about 62–4; *see also*  
 constitution  
 contingency of effect on 71–73, 75  
 distinguished from enabling conditions 98  
 dedicated to effect 80  
 descriptions of 68 n. 27  
 as enabling effect *see* enabler  
 as ennobling effect *see* ennobler  
 as enough for effect 32, 76–7, 98  
 essence of *see* causal essence  
 as events 70  
 and explanation 64, 83  
 latitudinal and longitudinal extent of 61  
 n. 8  
 and needs *see* needs  
 overdetermining effect *see*  
 overdetermination  
 overestimation of 73, 75  
 preemptive *see* preemption  
 as proportional to effect 77–81  
 as required for effect 32, 75–6, 94–5,  
 98  
 size and strength of 59–62  
 and switching 113–15  
 and truth-makers 301 n. 10  
 underestimation of 73  
 world- and effect-driven 80–83  
 Chisholm, Roderick 155  
 Clinton, Hillary 152  
 cogito 145, 202  
 coincidence 22–3, 46, 48  
 and categorical properties 25–6, 28–9,  
 65–7  
 and constitution 46  
 holding accidentally 67  
 and identity 29–30, 32  
 and intrinsicness 50–1, 52, 57  
 and ontology 78  
 and refinements 29  
 and strengthening 67  
 complete essence 16  
 complete profile 16  
 consistency-truth principle 202

- constitution 62–4, 68  
   and coincidence 46  
   constitutionalism 62–4  
   and copying 51–52  
   and essence 68  
   and intrinsicness 47  
   properties had constitutively 62, 68  
 constitutionalism *see* constitution  
 content  
   access to *see also* as-if 134–7, 150–1,  
     163–6, 228–9, 232, 235, 238  
   assertive 279, 283, 286, 289–90, 304, 306,  
     309–11  
   carving at the joints *see* carving content  
   conventional 211  
   literal 133–4, 139, 140–1, 160, 162–3,  
     168, 182, 197, 210–1  
   material content 229 n. 16  
   metaphorical 133–4, 162–3, 168, 233  
   multiple contents 187  
   of a game 131, 230–1  
   and platonism 152, 154  
   problem of real content *see* real content  
   propositional 136, 137, 164  
   representational 137, 165–6, 167  
   as sense *see also* sense 247  
 contingent identity 13, 22–9, 31  
   and essentialism 14–16  
 contingent properties 18  
   contingent existence 20  
   and contingent identity 23  
   distinguished from accidental 71 n. 33  
 copies 47, 51–52  
 count as true/false/gappy 289–91, 292,  
   300–302, 302–3, 304–6  
 counterfactual dependence 106–7, 108,  
   113–15  
 counterfactual objects 55–7  
 counterfactual properties 24, 48; *see also*  
   counterfactual relations  
 counterfactual relations 31  
 counterfactual worlds 53, 55  
 counterparts 40 n.19, 41 *see also* needs 104–5,  
   107, 112  
 covariation 33  
 Cruise, Tom 297, 309  
 cumulative essence 18–19, 29–30  
 cumulative property 18, 24 n. 15, 28, 29,  
   65–6, 71 n. 33  
  
 Davidson, Donald 61–64, 85 n. 68, 137,  
   155–6, 165, 167  
 de facto *see also* de facto dependence 33–4, 52,  
   71  
   de facto dependence 105–6, 108–9, 112–13,  
     114  
     and counterfactual dependence 106  
   de jure 33–5, 52  
   delayers 110–1  
   dependence modulo a condition 100–103,  
     110, 112, 113  
     and counterfactual dependence 106  
     and needs 104  
     and switching 114  
   Descartes, Rene 259  
   determinate/determinable relation *see*  
     property  
   determination 82, 87–90, 93–4  
   determinism 84  
   double effect 124, 125, 126  
   dualism 83–4, 86–7  
   duplication *see also* copies  
     and accompaniment 53 n. 41  
     categorical 68, 70  
     categorical duplicates 68, 70  
     immediate 52  
     intrinsic 194  
     and intrinsicness 42, 51–52  
     and naturalness 52  
     physical 207  
  
 effect  
   accidental 80  
   and categorical duplicates 68  
   causally guaranteed 86–7  
   cause as adequate for 72–5, 76  
   cause as commensurate to 59–60, 93,  
     94  
   cause as dedicated to 79–81  
   cause as enough for 76–7, 90–1, 98  
   cause as proportional to 77  
   cause as relevant to 91–92  
   as contingent on cause 71–72, 75  
   as dependent on effect *see* dependence  
     modulo a condition  
   effect-driven 81–83  
   enabler of *see* enabler  
   enabling conditions of 98  
   enfeebler of *see* enfeebler  
   ennobler of *see* ennobler  
   and mental causation 93–4  
   as needing its cause *see* needs  
   overdetermination of *see* overdetermination  
   and preemptive causes 100, 102, 108–9,  
     113  
   as requiring cause 75–6, 98  
   and size or strength of cause 61–62,  
     70

- effect (*cont.*)  
 and switching 113–15  
 threatened by cause *see* Stockholm syndrome
- Einstein, Albert 196
- empiricism  
 Frege's empiricism about structure 248–9  
 and Platonism 127
- enabler 99
- enfeebler 105–6, 109, 112, 113
- ennobler 99, 101, 105–6
- epiphenomenalism 83–6, 92, 93
- equivalence modulo 312
- essence *see also* essentialism, essential properties  
 of causes *see* causal essence  
 and causal powers 71, 72, 74, 77  
 and coincidence 25, 67  
 complete essence 16  
 and constitution 46, 68–70  
 and contingency 72, 75–6  
 cumulative essence 18–19, 29–30, 66  
 definition 19–22  
 difference between essences 21, 66, 89  
 and descriptions 68 n. 27  
 essentialization *see* property  
 and existence 20, 201  
 and extrinsic properties 47  
 and identity 16–17, 19, 29–30, 64  
 and identity of mental and physical 31–32, 86, 89–90, 94–5  
 inclusion relation among essences 17–19, 20; *see also* strengthening  
 and intrinsic properties 38, 40, 44, 45  
 and mereology 42–4  
 and necessity 204, 221  
 and parts 45–6  
 and proportionality of cause to effect 77  
 relation holding essentially 67  
 and restrictive properties 18
- essential properties 20, 21, 25, 38, 48, 51, 66–7, 68, 71, 200–1  
 absolutely essential 40–1  
 and categorical properties 65  
 and causally irrelevant properties 72, 75–6  
 and causation 59  
 and complete essence 16  
 and cumulative properties 17, 29  
 distinguished from necessary 71 n. 33  
 essential identity 16  
 and existence 20  
 and extrinsic properties 42–6, 47  
 and identity 14, 64  
 and intrinsic properties 43–4  
 and mind-body identity 86, 89–90  
 and necessity 204
- essentialism 14–16, 42, 64, 68–70
- essentialization 20–22, 28–9
- Etchemendy, John 147
- event  
 dedicated 80–1  
 mental 31, 84–90, 92 n. 81, 93–5; *see also* mental phenomena  
 physical 31, 84–5, 86, 89–90, 92 n. 81  
 proportional 77  
 switch 113–15
- exclusion argument 84–5, 87, 90–2
- exclusion principle *see* exclusion argument
- existence 46
- existence-claim 119, 120, 122, 142, 202
- extrinsic power 50
- extrinsic properties 33, 36, 38, 200  
 absolutely essential 40–1, 42–4  
 and essence 44–6, 47–8  
 and hypothetical properties 50, 51  
 and laws 49  
 and loneliness/accompaniment 34, 40, 53 n. 41
- Eyre, Jane 25 n. 16
- fallback scenario 103–5, 107, 108; *see also* needs
- fictionalism 179, 185, 190, 193–4, 197, 267  
 figurative fictionalism 191–92, 193  
 instrumentalist 179–80  
 meta-fictionalism 180–1  
 modal fictionalism 183  
 object fictionalism 181–5, 186, 187  
 property fictionalism 184–5  
 reflexive fictionalism 186–7, 187–9  
 relative reflexive fictionalism 187–9, 191  
 the Bomb 183–6
- Field, Harry 98, 118, 155, 180–1, 201 n. 2, 206–8, 224–9
- figuralism 191–94, 195–7, 208–9, 267
- Fine, Kit 43, 48 n. 33
- frameworks 119, 123, 126  
 adoption of 124–5, 127  
 and internal/external questions 121  
 and numbers 190  
 and pretense 128–9, 235
- Frege, Gottlob  
 arithmetic 222  
 Basic Law V 257  
 content-carving 246–51, 252  
 definition of numbers 183, 200 n. 1, 202, 212, 218  
 modality 253–4  
 names 286  
 presuppositions 272–3  
 rationalism 258–9

- fullness 25 n. 18, 28, 50–1  
 functionalism 31
- Goodmanian language 237–42  
 grasping content 246–7, 258
- Hale, Bob 205, 255, 258, 260  
 Hall, Ned 115  
 hasteners  
   as preempters 110–1  
   preempters as 111–12  
 Heck, Richard 258  
 Hills, David 138 n. 62, 235  
 holding fixed *see* dependence modulo a condition  
 Holmes, Sherlock 131, 231, 299, 303  
 Hume, David 59–60, 63–4, 71, 99  
 Hume's Principle 257, 267  
 hypothetical properties 24–5, 27, 31, 48, 49–50, 50–1, 65, 66  
   and causation 59, 70–1  
   and essential properties 68
- ideal objects 128  
 identity  
   and agreement of properties 22, 23, 25, 64  
   and complete essences 16–18  
   and complete profiles 16  
   of contents 246, 255  
   contingent *see* contingent identity  
   and essence 64–5, 67 n. 24, 200  
   indiscernibility of identicals 14  
   as intrinsic or extrinsic 41–42, 45, 50  
   of mental and physical *see* mind/body identity  
   necessary 13, 15, 27, 41  
   of numbers 185, 195  
   numerical 83  
   relation to coincidence 29–30  
   relative 31  
   as supervenient on other properties 17–8, 19  
 identity-like relations 15, 64  
 ideology/ontology tradeoff 135 n. 53, 155, 164 n. 24  
 inaccessible cardinals 242  
 indeterminism 94  
 indiscernibility  
   absolute 27  
   categorical 23, 66–7  
   and hypotheticality of identity 50  
   of identicals 14  
   indiscernible spheres 30  
   and provisionally categorical properties 26 n. 19
- inference to the best explanation *see* abduction  
 intrinsic properties 33, 34–6, 200, 259  
   accidental *see* problem of accidental  
   intrinsic  
   and absolutely essential properties 43–4  
   and categoricity 50  
   and constitution 47  
   definition of 39–40  
   and duplicates 51–52  
   and essence 44–6  
   generic and specific 48–50, 50–1  
   and identity 41–42, 50  
   and modal realism 53–7  
   and overlap 38–9  
   and presupposition failure 290  
 intrinsic/extrinsic distinction 50, 65 n. 22; *see also* properties
- Kaplan, David 277, 284  
 Kim, Jaegwon 34  
 Kripke, Saul  
   essential properties 14, 15 n. 4, 20 n. 12, 40  
   mind/brain identity 31, 86
- Langendoen, D. 283, 285  
 Langton, Rae 34–5, 42, 52  
 Lardner, Ring 295, 296  
 Leibniz, Gottfried Wilhelm 16, 17 n. 7  
 Leibniz's law 31, 246  
 Lewis, C.I. 151–52  
 Lewis, David  
   concrete worlds 55, 155  
   essentialism 74 n. 21  
   intrinsicness 34–5, 39, 41–42, 51, 52–3  
   natural sets 260  
   overlap 37–8, 55  
   preemption 99–100, 108  
   supervenience of truth on being 157  
 logically true<sub>cc</sub> 217–18  
 loneliness 34
- Maddy, Penelope 242  
 make-believe 120, 137, 140, 299; *see also* metaphor  
   and fictionalism 179, 187  
   and frameworks 128–9  
   and ontological commitment 129–30  
   and platonism 172–3  
   world-(prop)-oriented 131–33, 160, 231–35
- Marcus, Ruth 13  
 material biconditional 105  
 material object 40 n. 18, 121, 123, 170, 227 n. 13, 240 n. 30; *see also* concrete object  
 materialism 15 n. 4, 83 n. 61, 128 n. 34

- mathematical objects 158, 190, 207, 222, 225–6, 228–9, 230, 233, 242; *see also* numbers  
 detaching concrete content of statements about 229 n. 16  
 invention of 237  
 non-deductive usefulness of 226–7
- mathematical truth  
 absoluteness of 203–4, 208, 219  
 consistency-truth principle 202
- mathematics  
 development of 230
- mental events *see* mental phenomena
- mental phenomena 83–92  
 and causation 92–4
- mental properties  
 and causation 85–6  
 determined by physical properties 88–90  
 and functional properties 31
- mere-Cambridge power *see* extrinsic power
- metaphor 132–4, 139–41, 160–62, 166–9, 171, 232; *see also* representation  
 existential metaphor 135, 162, 163, 167, 232  
 felt-distance test for  
 metaphoricality 169–71  
 fifth grade of metaphorical involvement 141 n. 71  
 and figuralism 191  
 fourth grade of metaphorical involvement 138 n. 67  
 literal and metaphorical content of 133–4, 162–3, 233  
 literal/metaphorical distinction 120, 141 n. 73  
 and mathematics 234–7, 241–42  
 maybe-metaphor 169, 171  
 patient 138, 168  
 platonic 169, 171  
 pregnant 138, 168, 169  
 presentational force of 137, 165–6  
 presentationally essential 136–7  
 procedurally essential 137–8, 167  
 prophetic 138, 168  
 representationally essential 134–6, 139, 164  
 three grades of metaphorical involvement 138, 167–8  
 true/apt distinction 234–6
- Mill, JS 59–61, 63–4, 82, 286
- mind/body identity 31–32, 83–4, 85  
 and causation *see* causation  
 as a contingent identity 13 n. 2  
 and determination relation 87–90, 92  
 and multiple realizability 85, 86  
 and supervenience 88, 89  
 token identity 85, 86  
 type identity 86
- modal realism 36, 53–7
- models  
 assuming existence of 150–1, 154–5  
 believing in 197  
 existence of 145–7, 174–5  
 and objectivity 157–9  
 and ontological commitment 151–52  
 and ontology 147–8  
 real existence of 148–9
- monolithic mereologism 43–4, 45
- moot  
 existence questions 119, 143, 312  
 ontological mootness 297–8, 309
- Moran, Richard 136, 165
- necessary properties 13, 14, 19–20  
 distinguished from essential 71 n. 33  
 and identity 22 n. 13, 27
- needs 60–1, 76, 82  
 actual 104–5, 112, 114  
 artificial 103–7, 108, 109–10, 112, 114  
 cancelled 107–8, 111, 114–15  
 counterpart 104–5, 107, 112  
 and dependence 105–6  
 fallback 103–7, 108, 111, 112, 114, 115  
 and hasteners/delayers 111, 112  
 meets the same need as 104–5, 107; *see also* counterparts  
 puts in need of 100, 102, 103–4, 105, 106, 108–9
- Nixon, Richard 141, 142, 168, 171, 235
- nonactual worlds 24
- nonplatonic bridge principle 149
- numbers  
 intrinsic nature of 169–70, 201  
 as representational aids 186–8, 195–6, 218, 227–9, 230, 233–4  
 and theory revision 227 n. 13
- objective discourse *see* objectivity
- objectivity 157, 158–9, 194–5, 230
- O'Brien, Flann 170
- overdetermination 100, 109–10
- Parity Principle 257
- parts 5, 11, 37, 40, 43, 46, 53–4, 91, 311
- part/whole relation  
 and duplication 51  
 and intrinsicness 35–6, 47, 52, 53  
 mereological 43  
 among possible worlds 36–7, 38–40, 44–6, 53–7
- Peano arithmetic (PA) 202–3, 219, 237  
 AP 202–3, 219
- piece 45

- $\pi$ -free implication 287–291, 293, 301–304, 309, 311–13
- Plantinga, Alvin 145
- Plato 98–9
- platonian objects 146, 151–52, 154–6, 158–9, 171–72; *see also* abstract objects
- as-if interpretation of 174
- simulating 162
- platonism 150–58, 193–4, 221
- fregean 247, 266–7
- polarity problem 102, 106
- portion 46
- possibilization 20–22
- Potter, Michael 255, 260
- preconceived objects 7–8
- preemption 99–100, 108–9, 113; *see also* Stockholm syndrome
- and switching 113–15
- presupposition 269
- abstract existence 304, 305, 306, 308, 310–12
- concrete existence 304, 305, 306, 310
- fail-safe existential 292, 307–312
- failure 277, 283
- semantic 315 n. 5
- tests for 270–1
- uniqueness 274–5
- presupposition failure 269
- catastrophic 303, 314
- disruptive 271
- Donnellan and Stalnaker on 276–8
- Frege and Strawson on 273–6
- and ignore strategy 278–80
- non-catastrophic (NCPF) 269, 272–73, 288–9, 303, 306
- and restore strategy 280–82
- and say-more strategy 282–86
- of sentences that seem false 288, 289–92, 301
- of sentences that seem true 289–92, 302–3
- and truth conditions 273
- presuppositionalism 267
- pretended true 131–33, 182–3, 183–5, 187, 231, 232–3, 237, 303
- Prior, Arthur 155, 173 n. 45
- problem of accidental intrinsics 37–8, 41, 53
- proliferation problem 256, 258, 260
- properties; *see also* essence
- absolutely essential 40–1, 43; *see also* extrinsic properties
- accidental *see* accidental properties
- accidentalization of 20–22
- agreement and concurrence on 52
- and anti-realism 157, 171
- basic intrinsic 34
- categorical *see* categorical properties
- causal *see* causal properties
- classificatory 18 n. 8
- contingent *see* contingent properties
- and contingent identity 22–7, 28–9
- contrasted with attribute 20
- cumulative 18, 28, 29, 65–6; *see also* cumulative essence
- definite 297
- and determinate/determinable relation 87–90, 90–92
- disjunctive 34–5
- and essence 16–19
- essential 38, 65, 67, 68, 71; *see also* essence, causal properties
- essential *see* essential properties
- essentialization of 20–22, 28–9
- of event 90
- existence of 117, 146–7, 157, 159, 182; *see also* abstract objects, platonian objects
- extrinsic *see* extrinsic properties
- functional 31
- game-independent *see* make-believe
- hypothetical *see* hypothetical properties
- identity 17–8
- intrinsic *see* intrinsic properties
- kind 18, 65
- and manner of possession *see* constitution
- of mathematical objects *see* mathematical objects
- mental 31, 83–90
- model *see* property-model
- natural 34–5, 157
- necessary *see* necessary properties
- non-referential 309
- notion of 19–22
- of numbers *see* numbers
- phenomenal 31
- physical 85–90, 282
- possibilization of 20
- property fictionalism 184–5
- relational 200
- relative strength of 23
- restrictive 18
- satisfiable set of 20
- property-model 21
- downward-closed 23
- full 28
- maximal closed 28
- separable 30
- upward-closed 21
- Putnam, Hilary 87, 118, 180, 223 n. 6, 225
- quasi-assertion 180–1, 183–4, 186–7
- Quine, Willard Van Orman
- on commitment 117–18, 120, 160, 177–80, 242
- indispensability argument 118, 145, 147, 225, 273

- Quine, Willard Van Orman (*cont.*)  
 and internal/external distinction 119,  
 122–30  
 and metaphor 133, 135, 138–42, 164, 191  
 proxy functions 252  
 and quantified modal logic 62 n. 12  
 states of affairs 136 n. 54
- rationalism 147, 149, 249, 258–9  
 real content 180–1, 182–3, 187, 196, 210–1  
 logically true 219  
 of numerical statements 211–13  
 and objectual/assertional reality 190–1  
 problem of 179  
 of set theoretic statements 213–18  
 refinement 20–22, 26, 28–30  
 representation *see also* representational task  
 and content-carving 256, 259–60  
 material falsity of 258–60  
 and metaphor 132, 161, 166, 232  
 and necessity 218  
 representational aid 136, 164, 189, 210, 228  
 contrasted with thing represented 186  
 double role of 234  
 representational task  
 primary 259  
 secondary 259, 260  
 tertiary 260 n. 9  
 Rice, Hugh 111  
 Rosen, Gideon 148, 184, 221, 229 n. 15  
 Russell, Bertrand 200 n. 1, 248–9, 275 n. 13,  
 279, 299–300
- satisfaction of a sentence 254  
 say-hypothesis 159  
 Schaffer, Jonathan 113  
 Sellars, Wilfrid 237  
 sense 247, 250, 251–52, 253–4  
 sets  
 absoluteness of 203–4, 219–20  
 essential properties of 200  
 existence of 118, 146, 147, 151, 159, 177  
 real content of statements about 213–18  
 as representational aids 195  
 shmabstract objects 205, 206  
 Shoemaker, Sydney 49  
 Shope, Robert 282  
 simulated belief (belief per accidens) 196–7  
 slingshot argument 252  
 Smiley, Timothy 255, 260  
 Stalnaker, Robert 269 n. 2, 271, 272, 275  
 n. 12, 276–9, 282, 284–6, 288  
 state 22, 67  
 modal 29 n. 22
- Stockholm syndrome 102–3  
 Strawson, P.F. 271, 272, 273–80, 283–4,  
 299–301, 303, 306–7  
 strengthening *see also* refinement 17–19, 20,  
 65–7, 70, 77–8  
 and enoughness 76  
 and mind/body identity 86, 89–90  
 and proportionality 77–8  
 and world-drivenness 82 n. 58  
 Superman 259
- the Bomb *see* fictionalism  
 Thomson, Judy 46  
 Tombaugh, Clyde 273  
 truth *see also* truth values  
 absolute 208–9  
 according to *see* pretended true  
 analytic 123  
 and applicability 223, 224, 226  
 arithmetical *see* mathematics  
 on assumption 154, 158  
 and commitment 177  
 conceptual 152, 153–4  
 distinguished from agreement 190  
 and experience 149  
 factual 123  
 framework-independent (external) 126–8,  
 129  
 in fiction or pretense *see* pretended true  
 in virtue of meaning 123 n. 7  
 internal 129  
 literal 139, 140–1, 142, 148, 156, 160,  
 162–3, 168  
 logical 123, 183, 211–13, 214–8, 219,  
 233  
 metaphorical 133–4, 138, 140, 168, 235,  
 236  
 necessary 210  
 and ontology 149  
 and presupposition failure *see*  
 presupposition  
 relative 203–4, 207–8  
 supervenience on being 157  
 truth conditions 128, 140, 182–3  
 and unexpected (platonic) objects 146  
 truth profile 250–1  
 truth values 129, 157, 158, 159, 248–9  
 and content 250–1, 253–5, 278,  
 288–9  
 gaps in 306  
 and sense 251–52  
 and reference 203–12, 299–300  
 and presupposition failure *see*  
 presupposition

*Index*

323

- truth-falsity-maker 286–7, 289, 290, 292–3, 301  
unaboutness 142 n. 75  
uninvited company objection 255–6  
unique-sum principle 43–4
- Vallentyne, Peter 49  
value profile 251  
van Fraassen, Bas 180–1  
Vulcan 298, 299, 303
- Walton, Kendall 131–32, 170, 230–32  
Wiggins, David 18 n. 9, 31  
Wright, Crispin 145, 205, 267 n. 24
- Young, Neil 25 n. 16
- Zermelo numbers 173  
Zermelo-Fraenkel set theory (ZF) 204, 219–20  
FZ 204, 219–20

