NEW WORK FOR A THEORY OF UNIVERSALS

David Lewis

Introduction. D. M. Armstrong offers a theory of universals as the only adequate answer to a 'compulsory question' for systematic philosophy: the problem of One over Many.[1] I find this line of argument unpersuasvie. But I think there is more to be said for Armstrong's theory than he himself has said. For as I bear it in mind considering various topics in philosophy, I notice time and again that it offers solutions to my problems. Whatever we may think of the problem of One over Many, universals can earn their living doing other much-needed work.

I do not say that they are indispensable. The services they render could be matched using resources that are Nominalistic in letter, if perhaps not in spirit.[2] But neither do I hold any presumption against universals, to the effect that they are to be accepted only if we have no alternative. I therefore suspend judgement about universals themselves. I only insist that, one way or another, their work must be done.

I shall investigate the benefits of adding universals to my own usual ontology. That ontology, though Nominalistic, is in other respects generous. It consists of *possibilia* — particular, individual things, some of which comprise our actual world and others of which are unactualised[3] — together with the iterative hierarchy of classes built up from them. Thus I already have at my disposal a theory of properties as classes of *possibilia*. Properties, so understood, are not much like universals. Nor can they, unaided, take over the work of universals. Nevertheless they will figure importantly in what

---

[1] D. M. Armstrong, *Universals and Scientific Realism* (Cambridge University Press, 1978), henceforth cited as *'Universals'*; see also his "Against 'Ostrich' Nominalism: A Reply to Michael Devitt", *Pacific Philosophical Quarterly* 61 (1980) pp. 440-449.

[2] In this paper, I follow Armstrong's traditional terminology: 'universals' are repeatable entities, wholly present wherever a particular instantiates them; 'Nominalism' is the rejection of such entities. In the conflicting modern terminology of Harvard, classes count as 'universals' and 'Nominalism' is predominantly the rejection of classes. Confusion of the terminologies can result in grave misunderstanding; see W. V. Quine, 'Soft Impeachment Disowned', *Pacific Philosophical Quarterly* 61 (1980) pp. 450-451.

[3] Among 'things' I mean to include all the gerrymandered wholes and undemarcated parts admitted by the most permissive sort of mereology. Further, I include such physical objects as spatiotemporal regions and force fields, unless an eliminative reduction of them should prove desirable. Further, I include such nonphysical objects as gods and spooks, though not — I hope — as parts of the same world as us. Worlds themselves need no special treatment. They are things — big ones, for the most part.

follows, since for me they are part of the environment in which universals might operate.

The friend of universals may wonder whether they would be better employed not as an addition to my ontology of *possibilia* and classes, but rather as a replacement for parts of it. A fair question, and an urgent one; nevertheless, not a question considered in this paper.

In the next section, I shall sketch Armstrong's theory of universals, contrasting universals with properties understood as classes of *possibilia*. Then I shall say why I am unconvinced by the One over Many argument. Then I shall turn to my principal topic: how universals could help me in connection with such topics as duplication, supervenience, and divergent worlds; a minimal form of materialism; laws and causation; and the content of language and thought. Perhaps the list could be extended.

Universals and Properties. Language offers us several more or less interchangeable words: 'universal'; 'property', 'quality', 'attribute', 'feature', and 'characteristic'; 'type', 'kind', and 'sort'; and perhaps others. And philosophy offers us several conceptions of the entities that such words refer to. My purpose is not to fix on one of these conceptions; but rather to distinguish two (at opposite extremes) and contemplate helping myself to both. Therefore some regimentation of language is called for; I apologise for any inconvenience caused. Let me reserve the word 'universal' for those entities, if such there be, that mostly conform to Armstrong's account. And let me reserve the word 'property' for classes — any classes, but I have foremost in mind classes of things. To have a property is to be a member of the class.[4]

Why call them 'properties' as well as 'classes'? — Just to underline the fact that they need not be classes of *actual* things. The property of being a donkey, for instance, is the class of *all* the donkeys. This property belongs to — this class contains — not only the actual donkeys of this world we live in, but also all the unactualised, otherworldly donkeys.

Likewise I reserve the word 'relation' for arbitrary classes of ordered pairs, triples, ... Thus a relation among things is a property of 'tuples of things. Again, there is no restriction to actual things. Corresponding roughly to the division between properties and relations of things, we have the division between 'monadic' and 'polyadic' universals.

Universals and properties differ in two principal ways. The first difference concerns their instantiation. A universal is supposed to be wholly present wherever it is instantiated. It is a constituent part (though not a spatiotemporal part) of each particular that has it. A property, by contrast, is spread around. The property of being a donkey is partly present wherever there is a donkey,

---

[4] My conception of properties resembles the doctrine of Class Nominalism considered in *Universals*, I, pp. 28-43. But, strictly speaking, a Class Nominalist would be someone who claims to solve the One over Many problem simply by means of properties taken as classes, and that is far from my intention.

in this or any other world. Far from the property being part of the donkey, it is closer to the truth to say that the donkey is part of the property. But the precise truth, rather, is that the donkey is a member of the property.

Thus universals would unify reality (*Cf. Universals,* I, p. 109) in a way that properties do not. Things that share a universal have not just joined a single class. They literally have something in common. They are not entirely distinct. They overlap.

By occurring repeatedly, universals defy intuitive principles. But that is no damaging objection, since plainly the intuitions were made for particulars. For instance, call two entities *copresent* if both are wholly present at one position in space and time. We might intuit offhand that copresence is transitive. But it is not so, obviously, for universals. Suppose for the sake of argument that there are universals: round, silver, golden. Silver and round are copresent, for here is a silver coin; golden and round are copresent, for there is a gold coin; but silver and golden are not copresent. Likewise, if we add universals to an ontology of *possibilia,* for the relation of being part of the same possible world.[5] I and some otherworldly dragon are not worldmates; but I am a worldmate of the universal golden, and so is the dragon. Presumably I needed a mixed case involving both universals and particulars. For why should any two universals ever fail to be worldmates? Lacking such failures, the worldmate relation among universals alone is trivially transitive.

The second difference between universals and properties concerns their abundance. This is the difference that qualifies them for different work, and thereby gives rise to my interest in having universals and properties both.

A distinctive feature of Armstrong's theory is that universals are sparse. There are the universals that there must be to ground the objective resemblances and the causal powers of things, and there is no reason to believe in any more. All of the following alleged universals would be rejected:

not golden,	first examined before 2000 A.D.;
golden or wooden,	being identical,
metallic,	being alike in some respect,
self-identical,	being exactly alike,
owned by Fred,	being part of,
belonging to class C,	owning,
grue,	being paired with by some part in R

---

[5] If universals are to do the new work I have in store for them, they must be capable of repeated occurrence not only within a world but also across worlds. They would then be an exception to my usual principle — meant for particulars, of course — that nothing is wholly present as part of two different worlds. But I see no harm in that. If two worlds are said to overlap by having a coin in common, and if this coin is supposed to be wholly round in one world and wholly octagonal in the other, I stubbornly ask what shape it is, and insist that shape is not a relation to worlds. (See my 'Individuation by Acquaintance and by Stipulation', *Philosophical Review* 92 (1983), pp. 3-32.) I do not see any parallel objection if worlds are said to overlap by sharing a universal. What contingent, nonrelational property of the universal could we put in place of shape of the coin in raising the problem? I cannot think of any.

(where C and R are utterly miscellaneous classes). The guiding idea, roughly, is that the world's universals should comprise a minimal basis for characterising the world completely. Universals that do not contribute at all to this end are unwelcome, and so are universals that contribute only redundantly. A satisfactory inventory of universals is a non-linguistic counterpart of a primitive vocabulary for a language capable of describing the world exhaustively.

(That is rough: Armstrong does not dismiss redundant universals out of hand, as the spirit of his theory might seem to demand. Conjunctive universals — as it might be, golden-and-round — are accepted, though redundant; so are analysable structural universals. The reason is that if the world were infinitely complex, there might be no way to cut down to a minimal basis. The only alternative to redundancy might be inadequacy, and if so we had better tolerate redundancy. But the redundancy is mitigated by the fact that complex universals consist of their simpler — if perhaps not absolutely simple — constituents. They are not distinct entities. See *Universals,* II, pp. 30-42 and 67-71.)

It is quite otherwise with properties. Any class of things, be it ever so gerrymandered and miscellaneous and indescribable in thought and language, and be it ever so superfluous in characterising the world, is nevertheless a property. So there are properties in immense abundance. (If the number of things, actual and otherwise, is beth-2, an estimate I regard as more likely low than high, then the number of properties of things is beth-3. And that is a big infinity indeed, except to students of the outer reaches of set theory.) There are so many properties that those specifiable in English, or in the brain's language of synaptic interconnections and neural spikes, could be only an infinitesimal minority.

Because properties are so abundant, they are undiscriminating. Any two things share infinitely many properties, and fail to share infinitely many others. That is so whether the two things are perfect duplicates or utterly dissimilar. Thus properties do nothing to capture facts of resemblance. That is work more suited to the sparse universals. Likewise, properties do nothing to capture the causal powers of things. Almost all properties are causally irrelevant, and there is nothing to make the relevant ones stand out from the crowd. Properties carve reality at the joints — and everywhere else as well. If it's distinctions we want, too much structure is no better than none.

It would be otherwise if we had not only the countless throng of all properties, but also an élite minority of special properties. Call these the *natural* properties.[6] If we had properties and universals both, the universals

---

[6]   See *Universals,* I, pp. 38-41; Anthony Quinton, 'Properties and Classes', *Proceedings of the Aristotelian Society* 48 (1957) pp. 33-58; and W. V. Quine, 'Natural Kinds', in his *Ontological Relativity* (Columbia University Press, 1969). See also George Bealer, *Quality and Concept* (Oxford University Press, 1982), especially pp. 9-10 and 177-187. Like me, Bealer favours an inegalitarian twofold conception of properties: there are abundant 'concepts' and sparse 'qualities', and the latter are the ones that 'determine the logical, causal, and phenomenal order of reality'. (p. 10) Despite this point of agreement, however, Bealer's views and mine differ in many ways.

could serve to pick out the natural properties. Afterwards the universals could retire if they liked, and leave their jobs to the natural properties. Natural properties would be the ones whose sharing makes for resemblance, and the ones relevant to causal powers. Most simply, we could call a property *perfectly* natural if its members are all and only those things that share some one universal. But also we would have other less-than-perfectly natural properties, made so by families of suitable related universals.[7] Thus we might have an imperfectly natural property of being metallic, even if we had no such single universal as metallic, in virtue of a close-knit family of genuine universals one or another of which is instantiated by any metallic thing. These imperfectly natural properties would be natural to varying degrees.

Let us say that an *adequate* theory of properties is one that recognises an objective difference between natural and unnatural properties; preferably, a difference that admits of degree. A combined theory of properties and universals is one sort of adequate theory of properties.

But not the only sort. A Nominalistic theory of properties could achieve adequacy by other means. Instead of employing universals it could draw primitive distinctions among particulars. Most simply, a Nominalist could take it as a primitive fact that some classes of things are perfectly natural properties; others are less-than-perfectly natural to various degrees; and most are not at all natural. Such a Nominalist takes 'natural' as a primitive predicate, and offers no analysis of what he means in predicating it of classes. His intention is to select the very same classes as natural properties that the user of universals would select. But he regards the universals as idle machinery, fictitiously superimposed on the primative objective difference between the natural properties and the others.[8]

Alternatively, a Nominalist in pursuit of adequacy might prefer to rest with primitive objective resemblance among things. (He might not think that 'natural' was a very natural primitive, perhaps because it is to be predicated of classes.) Then he could undertake to define natural properties in terms of the mutual resemblance of their members and the failure of resemblance between their members and their non-members. Unfortunately, the project meets with well-known technical difficulties. These can be solved, but at a daunting price in complexity and artificiality of our primitive. We cannot get by with the familiar dyadic 'resembles'. Instead we need a predicate of resemblance that is both contrastive and variably polyadic. Something like

$x_1, x_2, \ldots$. resemble one another and do not likewise resemble any of y_1, $y_2, \ldots$

---

[7] Here I assume that some solution to the problem of resemblance of universals is possible, perhaps along the lines suggested by Armstrong in *Universals,* II, pp. 48-52 and 101-131; and that such a solution could be carried over into a theory of resemblance of perfectly natural properties, even if we take naturalness of properties as primitive.

[8] This is the Moderate Class Nominalism considered in *Universals,* I, pp. 38-41. It is akin to the view of Quinton, *op. cit.;* but plus the unactualised members of the natural classes, and minus any hint that 'natural' could receive a psychologistic analysis.

(where the strings of variables may be infinite, even uncountable) must be taken as understood without further analysis.[9] If adequate Nominalism requires us to choose between this and a primitive predicate of classes, we might well wonder whether the game is worth the candle. I only say we might wonder; I know of no consideration that seems to me decisive.

At this point, you may see very well why it could be a good idea to believe in universals as well as properties; but you may see no point in having properties as well as universals. But properties have work of their own, and universals are ill-suited to do the work of properties.

It is properties that we need, sometimes natural and sometimes not, to provide an adequate supply of semantic values for linguistic expressions. Consider such sentences as these:

(1)   Red resembles orange more than it resembles blue.
(2)   Red is a colour.
(3)   Humility is a virtue.
(4)   Redness is a sign of ripeness.

Prima facie, these sentences contain names that cannot be taken to denote particular, individual things. What is the semantic role of these words? If we are to do compositional semantics in the way that is best developed, we need entities to assign as semantic values to these words, entities that will encode their semantic roles. Perhaps sometimes we might find paraphrases that will absolve us from the need to subject the original sentence to semantic

---

[9]   Such a theory is a form of Resemblance Nominalism, in Armstrong's classification, but it is unlike the form that he principally considers. See *Universals,* I, pp. 44-63. For discussions of the problem of defining natural classes in terms of resemblance, and of the trickery that proves useful in solving this problem, see Nelson Goodman, *The Structure of Appearance* (Harvard University Press, 1951), Chapters IV-VI; W. V. Quine, 'Natural Kinds'; and Adam Morton, 'Complex Individuals and Multigrade Relations', *Noûs* 9 (1975) pp. 309-318.

To get from primitive resemblance to perfectly natural properties, I have in mind a definition as follows. We begin with R as our contrastive and variably polyadic primitive. We want it to turn out that $x_1,x_2, \ldots Ry_1,y_2, \ldots$ iff some perfectly natural property is shared by all of $x_1, x_2, \ldots$ but by none of $y_1, y_2, \ldots$. We want to define N, another variably polyadic predicate, so that it will turn out that $Nx_1,x_2, \ldots$ iff $x_1, x_2, \ldots$ are all and only the members of some perfectly natural property. Again we must allow for, and expect, the case where there are infinitely many x's. We define $Nx_1,x_2, \ldots$ as:
$$\exists y_1,y_2, \ldots \forall z\ (z,x_1,x_2, \ldots Ry_1,y_2, \ldots \equiv z = x_1 \vee z = x_2 \vee \ldots).$$
Then we finish the job by defining a perfectly natural property as a class such that, if $x_1, x_2, \ldots$ are all and only its members, then $Nx_1, x_2, \ldots$.

We might have taken N as primitive instead of R. But would that have been significantly different, given the interdefinability of the two? On the other hand, taking N as primitive also seems not significantly different from taking perfect naturalness of classes as primitive. It is only a difference between speaking in the plural of individuals and speaking in the singular of their classes, and that seems no real difference. Is plural talk a disguised form of class talk? Or *vice versa*? (See the discussion in *Universals,* I, pp. 32-34; also Max Black, 'The Elusiveness of Sets', *Review of Metaphysics* 24 (1971) pp. 614-636; Eric Stenius, 'Sets', *Synthese* 27 (1974), pp. 161-188; and Kurt Gödel, 'Russell's Mathematical Logic', in P. A. Schilpp, ed., *The Philosophy of Bertrand Russell* (Cambridge University Press, 1944).) At any rate, it is not at all clear to me that Moderate Class Nominalism and Resemblance Nominalism in its present form are two different theories, as opposed to a single theory presented in different styles.

analysis. That is the case with (1), for instance.[10] But even if such paraphrases sometimes exist — even if they *always* exist, which seems unlikely — they work piecemeal and frustrate any systematic approach to semantics.

Armstrong takes it that such sentences provide a subsidiary argument for universals, independent of his main argument from the One over Many problem. (*Universals,* I, pp. 58-63; also "Against 'Ostrich' Nominalism".[11]) I quite agree that we have here an argument for something. But not for universals as opposed to properties. Properties can serve as the requisite semantic values. Indeed, properties are much better suited to the job than universals are. That is plain even from the examples considered. It is unlikely that there are any such genuine universals as the colours (especially determinable colours, like red, rather than determinate shades), or ripeness, or humility. Armstrong agrees (*Universals,* I, p. 61) that he cannot take (1)-(4) as straightforwardly making reference to universals. He must first subject them to paraphrase. Even if there always is a paraphrase that does refer to, or quantify over, genuine universals, still the need for paraphrase is a threat to systematic sematics. The problem arises exactly because universals are sparse. There is no corresponding objection if we take the requisite semantic values as properties.

Other sentences make my point more dramatically.

(5)   Grueness does not make for resemblance among all its instances.
(6)   What is common to all who suffer pain is being in some or another state that occupies the pain role, presumably not the same state in all cases.

The point is not that these sentences are true — though they are — but that they require semantic analysis. (It is irrelevant that they are not ordinary language.) A universal of grueness would be anathema; as would a universal such that, necessarily, one has it if he is in some state or other that occupies

---

[10] In virtue of the close resemblance of red and orange, it is possible for a red thing to resemble an orange one very closely; it is not possible for a red thing to resemble a blue one quite so closely. Given our ontology of *possibilia,* all possibilities are realised. So we could paraphase (1) by

(1')   Some red thing resembles some orange thing more than any red thing resembles any blue thing.

so long as it is understood that the things in question needn't be part of our world, or of any one world. Or if we did not wish to speak of unactualised things, but we were willing to take ordinary-language modal idioms as primitive, we could instead give the paraphrase:

(1'')   A red thing can resemble an orange thing more closely than a red thing can resemble a blue thing.

It is necessary to use the ordinary-language idioms, or some adequate formalisation of them, rather than standard modal logic. You cannot express (1'') in modal logic (excluding an enriched modal logic that would defeat the point of the paraphrase by quantifying over degrees of resemblance or whatnot) because you cannot express cross-world relations, and in particular cannot express the needed cross-world comparison of similarity.

[11] He derives the argument, and a second semantic argument to be considered shortly, from Arthur Pap, 'Nominalism, Empiricism, and Universals: I', *Philosophical Quarterly* 9 (1959) pp. 330-340, and F. C. Jackson, 'Statements about Universals', *Mind* 86 (1977) pp. 427-429.

the pain role in his case.[12] But the corresponding properties are no problem.

Indeed, we have a comprehension schema applying to any predicate phrase whatever, however complicated. (Let it even be infinitely long; let it even include imaginary names for entities we haven't really named.) Let x range over things, P over properties (classes) of things. Then:

$$\exists_1 P \square \forall x \ (x \text{ has } P \equiv \emptyset x).$$

We could appropriately call this 'the property of $\emptyset$-ing' in those cases where the predicate phrase is short enough to form a gerund, and take this property to be the semantic value of the gerund. Contrast this with the very different relationship of universals and predicates set forth in *Universals*, II, pp. 7-59.

Consider also those sentences which *prima facie* involve second-order quantification. From *Universals*, I, p. 62, and "Against 'Ostrich' Nominalism" we have these.

(7) He has the same virtues as his father.
(8) The dresses were of the same colour.
(9) There are undiscovered fundamental physical properties.
(10) Acquired characteristics are never inherited.
(11) Some zoological species are cross-fertile.

Prima facie, we are quantifying either over properties or over universals. Again, paraphrases might defeat that presumption, but in a piecemeal way that threatens systematic semantics. In each case, properties could serve as the values of the variables of quantification. Only in case (9) could universals serve equally well. To treat the other cases, not to mention

(12) Some characteristics, such as the colours, are more disjunctive than they seem.

as quantifications over universals, we would again have to resort to some preliminary paraphrase. (Armstrong again agrees: *Universals*, I, p. 63.) This second semantic argument, like the first, adduces work for which properties are better qualified than universals.

Which is not to deny that a partnership might do better still. Let it be granted that we are dealing with quantifications over properties. Still, these quantifications − like most of our quantifications − may be tacitly or explicitly restricted. In particular, they usually are restricted to natural properties. Not to perfectly natural properties that correspond to single universals, except in special cases like (9), but to properties that are at least somewhat more natural than the great majority of the utterly miscellaneous. That is so for all our examples, even (12). Then even though we quantify over properties, we still need either universals or the resources of an adequate Nominalism in order to say which of the properties we mostly quantify over.

---

[12] Or better, in the case of creatures of his kind. See my 'Mad Pain and Martian Pain', in Ned Block, ed., *Readings in Philosophy of Psychology,* I (Harvard University Press, 1980).

I also think that it is properties that we need in characterising the content of our intentional attitudes. I believe, or I desire, that I live in one of the worlds in a certain class, rather than any world outside that class. This class of worlds is a property had by worlds. I believe, or I desire, that my world has that property. (The class of worlds also may be called a *proposition,* in one of the legitimate senses of that word, and my 'propositional attitude' of belief or desire has this proposition as its 'object'.) More generally, subsuming the previous case, I believe or I desire that I myself belong to a certain class of *possibilia.* I ascribe a certain property to myself, or I want to have it. Or I might ascribe a property to something else, or even to myself, under a relation of acquaintance I bear to that thing.[13] Surely the properties that give the content of attitudes in these ways cannot be relied on to be perfectly natural, hence cannot be replaced by universals. It is interesting to ask whether there is any lower limit to their naturalness (see the final section of this paper), but surely no very exacting standard is possible. Here again properties are right for the job, universals are not.

One Over Many. Armstrong's main argument for universals is the 'One over Many'. It is because I find this argument unconvincing that I am investigating alternative reasons to accept a theory of universals.

Here is a concise statement of the argument, taken by condensation from "Against 'Ostrich' Nominalism", pp. 440-441. A very similar statement could have been drawn from the opening pages of *Universals.*

> I would wish to start by saying that many different particulars can all have what appears to be the same nature and draw the conclusion that, as a result, there is a *prima facie* case for postulating universals. We are continually talking about different things having the same property or quality, being of the same sort or kind, having the same nature, and so on. Philosophers draw the distinction between sameness of token and sameness of type. But they are only making explicit a distinction which ordinary language (and so, ordinary thought) perfectly recognises. I suggest that the fact of sameness of type is a Moorean fact: one of the many facts which even philosophers should not deny, whatever philosophical account or analysis they give of such facts. Any comprehensive philosophy must try to give some account of Moorean facts. They constitute the compulsory questions in the philosophical examination paper.

From this point of departure, Armstrong makes his case by criticising rival attempts to answer the compulsory question, and by rejecting views that decline to answer it at all.

Still more concisely, the One over Many problem is presented as the problem of giving some account of Moorean facts of apparent sameness of type. Thus understood, I agree that the question is compulsory; I agree that

---

[13] See my 'Attitudes *De Dicto* and *De Se*', *Philosophical Review* 88 (1979) pp. 513-543; and 'Individuation by Acquaintance and by Stipulation'.

Armstrong's postulation of shared universals answers it; but I think that an adequate Nominalism also answers it.

An effort at systematic philosophy must indeed give an account of any purported fact. There are three ways to give an account. (1) 'I deny it' – this earns a failing mark if the fact is really Moorean. (2) 'I analyse it thus' – this is Armstrong's response to the facts of apparent sameness of type. Or (3) 'I accept it as primitive'. Not every *account* is an *analysis*! A system that takes certain Moorean facts as primitive, as unanalysed, cannot be accused of failing to make a place for them. It neither shirks the compulsory question nor answers it by denial. It does give an account.

An adequate Nominalism, of course, is a theory that takes Moorean facts of apparent sameness of type as primitive. It predicates mutual resemblance of the things which are apparently of the same type; or it predicates naturalness of some property that they all share, *i.e.* that has them all as members; and it declines to analyse these predications any further. That is why the problem of One over Many, rightly understood, does not provide more than a *prima facie* reason to postulate universals. Universals afford one solution, but there are others.

I fear that the problem does not remain rightly understood. Early in *Universals* it undergoes an unfortunate double transformation. In the course of a few pages (*Universals*, I, pp. 11-16) the legitimate demand for an account of Moorean facts of apparent sameness of type turns into a demand for an analysis of predication in general. The *analysandum* becomes the schema '*a* has the property *F*'. The turning point takes only two sentences (p. 12):

> How is [the Nominalist] to account for the apparent (if usually partial) identity of numerically different particulars? How can two different things both be white or both be on a table?

And very soon (pp. 16-17) those who 'refuse to countenance universals but who at the same time see no need for any reductive analyses [of the schema of predication]', those according to whom 'there are no universals but the proposition that *a* is *F* is perfectly all right as it is' stand accused of dodging the compulsory question.

When the demand for an account – for a place in one's system – turned into a demand for an analysis, then I say that the question ceased to be compulsory. And when the *analysandum* switched, from Moorean facts of apparent sameness of type to predication generally, then I say that the question ceased to be answerable at all. The transformed problem of One over Many deserves our neglect. The ostrich that will not look at it is a wise bird indeed.

Despite his words, I do not think that Armstrong really means to demand, either from Nominalists or from himself, a *fully* general analysis of predication. For none is so ready as he to insist that not just any shared predicate makes for even apparent sameness of type. (That is what gives his theory its distinctive interest and merit.) It would be better to put the transformed problem thus: one way or another, all predication is to be

analysed. Some predications are to be analysed away in terms of others. Here we have one-off analyses for specific predicates — as it might be, for 'grue'. But all those predications that remain, after the one-off analyses are finished, are to be analysed wholesale by means of a general analysis of the schema '*a* has property *F*'.

There is to be no unanalysed predication. Time and again, Armstrong wields this requirement against rival theories. One theory after another falls victim to the 'relation regress': in the course of analysing other predications, the theory has resort to a new predicate that cannot, on pain of circularity, be analysed along with the rest. So falls Class Nominalism (including the version with primitive naturalness that I deem adequate): it employs predications of class membership, which predications it cannot without circularity analyse in terms of class membership. So falls Resemblance Nominalism: it fails to analyse predications of resemblance. So fall various other, less deserving Nominalisms. And so fall rival forms of Realism, for instance Transcendent, Platonic Realism: this time, predications of participation evade analysis. Specific theories meet other, specific objections; suffice it to say that I think these inconclusive against the two Nominalisms that I called adequate. But the clincher, the one argument that recurs throughout the many refutations, is the relation regress. And this amounts to the objection that the theory under attack does not achieve its presumed aim of doing away with all unanalysed predication and therefore fails to solve the transformed problem of One over Many.

Doing away with all unanalysed predication is an unattainable aim, and so an unreasonable aim. No theory is to be faulted for failing to achieve it. For how could there be a theory that names entities, or quantifies over them, in the course of its sentences, and yet altogether avoids primitive predication? Artificial tricks aside,[14] the thing cannot be done.

What's true is that a theory may be faulted for its overabundant primitive predications, or for unduly mysterious ones, or for unduly complicated ones. These are not fatal faults, however. They are to be counted against a theory, along with its faults of overly generous ontology or of disagreement with less-than-Moorean commonsensical opinions. Rival philosophical theories have their prices, which we seek to measure. But it's all too clear that for philosophers, at least, there ain't no such thing as a free lunch.

How does Armstrong himself do without primitive predication? — He doesn't. Consider the predicate 'instantiates' (or 'has'), as in 'particular *a*

---

[14] Let *S* be the syntactic category of sentences, let *N* be the category of names, and for any categories *x* and *y*, let *x/y* be the category of expressions that attach to *y*-expressions to make *x*-expressions. Predicates, then, are category *S/N*. (Or (*S/N*)/*N* for two-place predicates, and so on.) To embed names (or variables in the category of names) into sentences without primitive predication, take any category *q* which is neither *S* nor *N*, nor *S/N*, and let there be primitives of categories *Q/N* and *S/Q*. Or take Q_1 and Q_2, different from *S* and *N* and *S/N* and each other, and let the primitives be of categories Q_1/N, Q_2/Q_1, and S/Q_2. Or . . . . I cannot see how this trickery could be a genuine alternative to, rather than a disguise for, primitive predication.

instantiates universal *F*' or 'this electron has unit charge'. No one-off analysis applies to this specific predicate. 'Such identity in nature [as results from the having of one universal in many particulars] is literally inexplicable, in the sense that it cannot be further explained.' (*Universals,* I, p. 109.) Neither do predications of 'instantiates' fall under Armstrong's general analysis of (otherwise unanalysed) predication. His is a *non-relational* Realism: he declines, with good reason, to postulate a dyadic universal of instantiation to bind particulars to their universals. (And if he did, it would only postpone the need for primitive predication.) So let all who have felt the bite of Armstrong's relation regress rise up and cry *'Tu quoque!'* And let us mark well that Armstrong is prepared to give *one* predicate 'what has been said to be the privilege of the harlot: power without responsibility. The predicate is informative, it makes a vital contribution to telling us what is the case, the world is different if it is different, yet ontologically it is supposed not to commit us. Nice work: if you can get it.' (Compare Armstrong on Quine's treatment of predication, "Against 'Ostrich' Nominalism", p. 443.)

Let us dump the project of getting rid of primitive predication, and return to the sensible — though not compulsory — project of analysing Moorean facts of apparent sameness of type. Now does the relation regress serve Armstrong better? I think not. It does make better sense within the more sensible project, but it still bites Armstrong and his rivals with equal force. Let the Nominalist say 'These donkeys resemble each other, so likewise do those stars, and there analysis ends.' Let the Platonist say 'This statue participates in the Form of beauty, likewise that lecture participates in the Form of truth, and there analysis ends.' Let Armstrong say 'This electron instantiates unit charge, likewise that proton instantiates tripartiteness, and there analysis ends.' It is possible to complain in each case that a fact of sameness of type has gone unanalysed, the types being respectively resemblance, participation, and instantiation. But it is far from evident that the alleged facts are Moorean, and still less evident that the first two are more Moorean than the third. None of them are remotely the equals of the genuine Moorean fact that, in some sense, different lumps of gold are the same in kind.

Michael Devitt has denounced the One over Many problem as a mirage better left unseen.[15] I have found Devitt's discussion instructive and I agree with much of what he says. But Devitt has joined Armstrong in transforming the One over Many problem. He takes it to be the problem of analysing the schema

a and *b* have the same property (are of the same type), *F*-ness

otherwise than by means of a one-off analysis for some specific *F*. To that problem it is fair to answer as he does that

[15] " 'Ostrich Nominalism' or 'Mirage Realism'?", *Pacific Philosophical Quarterly* 61 (1980) pp. 433-439. Devitt speaks on behalf of Quine as well as himself; Quine indicates agreement with Devitt in 'Soft Impeachment Disowned'.

a is F; b is F

is analysis enough, once we give over the aim of doing without primitive predication. But Devitt has set himself too easy a problem. If we attend to the modest, untransformed One over Many problem, which is no mirage, we will ask about a different *analysandum*:

a and *b* have some common property (are somehow of the same type)

in which it is not said what *a* and *b* have in common. This less definite *analysandum* is not covered by what Devitt has said. If we take a clearly Moorean case, he owes us an account: either an analysis or an overt resort to primitive predication of resemblance.

Duplication, Supervenience, and Divergent Worlds. Henceforth I shall speak only of my need for the distinction between natural and unnatural, or more and less natural, properties. It is to be understood that the work I have in store for an adequately discriminatory theory of properties might be new work for a theory of universals, or it might instead be work for the resources of an adequate Nominalism.

I begin with the problem of analysing duplication. We are familiar with cases of approximate duplication, *e.g.* when we use copying machines. And we understand that if these machines were more perfect than they are, the copies they made would be perfect duplicates of the original. Copy and original would be alike in size and shape and chemical composition of the ink marks and the paper, alike in temperature and magnetic alignment and electrostatic charge, alike even in the exact arrangement of their electrons and quarks. Such duplicates would be exactly alike, we say. They would match perfectly, they would be qualitatively identical, they would be indiscernible.

But they would not have exactly the same properties, in my sense of the word. As in the case of any two things, countless class boundaries would divide them. Intrinsically, leaving out their relations to the rest of the world, they would be just alike. But they would occupy different spatio-temporal positions; and they might have different owners, be first examined in different centuries, and so on.

So if we wish to analyse duplication in terms of shared properties, it seems that we must first distinguish the *intrinsic* (or 'internal') properties from the *extrinsic* (or 'external' or 'relational') properties. Then we may say that two things are duplicates iff they have precisely the same intrinsic properties, however much their extrinsic properties might differ. But our new problem of dividing the properties into intrinsic and extrinsic is no easier than our original problem of analysing duplication. In fact, the two problems are joined in a tight little circle of interdefinability. Duplication is a matter of sharing intrinsic properties; intrinsic properties are just those properties that never differ between duplicates. Property P is intrinsic iff, for any two duplicate things, not necessarily from the same world, either both have P

or neither does. *P* is extrinsic iff there is some such pair of duplicates of which one has *P* and the other lacks *P*.[16]

If we relied on our physical theory to be accurate and exhaustive, we might think to define duplication in physical terms. We believe that duplicates must be alike in the arrangement of their electrons and quarks — why not put this forward as a definition? But such a 'definition' is no analysis. It presupposes the physics of our actual world; however physics is contingent and known *a posteriori*. The definition does not apply to duplication at possible worlds where physics is different, or to duplication between worlds that differ in their physics. Nor does it capture what those ignorant of physics mean when they speak — as they do — of duplication.

The proper course, I suggest, is to analyse duplication in terms of shared properties; but to begin not with the intrinsic properties but rather with natural properties. Two things are qualitative duplicates if they have exactly the same perfectly natural properties.[17]

Physics is relevant because it aspires to give an inventory of natural

---

[16] Given duplication, we can also subdivide the extrinsic properties, distinguishing pure cases from various mixtures of extrinsic and intrinsic. Partition the things, of this and other worlds, into equivalence classes under the relation of duplication. A property may divide an equivalence class, may include it, or may exclude it. A property *P* is extrinsic, as we said, if it divides some of the classes. We have four subcases. (1) *P* divides every class; then we may call *P purely extrinsic*. (2) *P* divides some classes, includes some, and excludes none; then *P* is the disjunction of an intrinsic property and a purely extrinsic property. (3) *P* divides some, excludes some, and includes none; then *P* is the conjunction of an intrinsic property and a purely extrinsic property. (4) *P* divides some, includes some, and excludes some; then *P* is the conjunction of an intrinsic property and an impurely extrinsic property of the sort considered in the second case, or equivalently is the disjunction of an intrinsic property and an impurely extrinsic property of the sort considered in the third case.

We can also classify relations as intrinsic or extrinsic, but in two different ways. Take a dyadic relation, *i.e.* a class or ordered pairs. Call the relation *intrinsic to its relata* iff, whenever *a* and *a'* are duplicates (or identical) and *b* and *b'* are duplicates (or identical), then both or neither of the pairs $<a,b>$ and $<a',b'>$ stand in the relation. Call the relation *intrinsic to its pairs* iff, whenever the pairs $<a,b>$ and $<a',b'>$ themselves are duplicates, then both or neither of them stand in the relation. In the second case, a stronger requirement is imposed on the pairs. For instance they might fail to be duplicate pairs because the distance between *a* and *b* differs from the distance between *a'* and *b'*, even though *a* and *a'* are duplicates and *b* and *b'* are duplicates. In traditional terminology, 'internal relations' are intrinsic to their *relata*; 'external relations' are intrinsic to their pairs but not to their *relata*; and relations extrinsic even to their pairs, such as the relation of belonging to the same owner, get left out of the classification altogether.

Our definition of intrinsic properties in terms of duplication closely resembles the definition of 'differential properties' given by Michael Slote in 'Some Thoughts on Goodman's Riddle', *Analysis* 27 (1967) pp. 128-132, and in *Reason and Scepticism* (George Allen & Unwin, 1970). But where I quantify over *possibilia,* Slote applies modality to ordinary, presumably actualist, quantifiers. That makes a difference. An extrinsic property might differ between duplicates, but only when the duplicates inhabit different worlds; then Slote would count the property as differential. An example is the property of being a sphere that inhabits a world where there are pigs or a cube that inhabits a world without pigs.

See my 'Extrinsic Properties', *Philosophical Studies* 44 (1983) for further discussion of the circle from duplication to intrinsicness and back.

[17] Likewise $<a,b>$ and $<a',b'>$ and duplicate pairs iff *a* and *a'* have exactly the same perfectly natural properties, and so do *b* and *b'*, and also the perfectly natural relations between *a* and *b* are exactly the same as those between *a'* and *b'*.

properties — not a complete inventory, perhaps, but a complete enough inventory to account for duplication among actual things. If physics succeeds in this, then duplication within our world amounts to sameness of physical description. But the natural properties themselves are what matter, not the theory that tells us what they are. If Materialism were false and physics an utter failure, as is the case at some deplorable worlds, there would still be duplication in virtue of shared natural properties.

On my analysis, all perfectly natural properties come out intrinsic. That seems right. The converse is not true. Intrinsic properties may be disjunctive and miscellaneous and unnatural, so long as they never differ between duplicates. The perfectly natural properties comprise a basis for the intrinsic properties; but arbitrary Boolean compounds of them, however unnatural, are still intrinsic. Hence if we adopt the sort of adequate Nominalism that draws a primitive distinction between natural and unnatural properties, that is not the same thing as drawing a primitive distinction between intrinsic and extrinsic properties. The former distinction yields the latter, but not *vice versa*.

Likewise if we adopt the sort of adequate Nominalism that begins with a suitable relation of partial resemblance, that is not the same thing as taking duplication itself as primitive. Again, the former yields the latter, but not *vice versa*.

If instead we reject Nominalism, and we take the perfectly natural properties to be those that correspond to universals (in the sense that the members of the property are exactly those things that instantiate the universal), then all the properties that correspond to universals are intrinsic. So are all the Boolean compounds — disjunctions, negations, etc. — of properties that correspond to universals. The universals themselves are intrinsic *ex officio,* so to speak.

But here I must confess that the theory of universals for which I offer new work cannot be exactly Armstrong's theory. For it must reject extrinsic universals; whereas Armstrong admits them, although not as irreducible. (See *Universals,* II, pp. 78-79.) I think he would be better off without them, given his own aims. (1) They subvert the desired connection between sharing of universals and Moorean facts of partial or total sameness of nature. Admittedly, there is such a thing as resemblance in extrinsic respects: things can be alike in the roles they play *vis-à-vis* other things, or in the origins they spring from. But such resemblances are not what we mean when we say of two things that they are of the same kind, or have the same nature. (2) They subvert the desired immanence of universals: if something instantiates an extrinsic universal, that is not a fact just about that thing. (3) They are not needed for Armstrong's theory of laws of nature; any supposed law connecting extrinsic universals of things can be equivalently replaced by a law connecting intrinsic structures of larger systems that have those things as parts.

Thus I am content to say that if there are universals, intrinsic duplicates are things having exactly the same universals. We need not say '. . . exactly the same *intrinsic* universals' because we should not believe in any other kind.

Not only is duplication of interest in its own right; it also is needed in dealing with other topics in metaphysics. Hence such topics create a derived need for natural properties. I shall consider two topics where I find need to speak of duplication: supervenience and divergent worlds.

*   *   *

First, supervenience. A supervenience thesis is a denial of independent variation. Given an ontology of *possibilia,* we can formulate such theses in terms of differences between possible individuals or worlds. To say that so-and-so supervenes on such-and-such is to say that there can be no difference in respect of so-and-so without difference in respect of such-and-such. Beauty of statues supervenes on their shape, size, and colour, for instance, if no two statures, in the same or different worlds, ever differ in beauty without also differing in shape or size or colour.[18]

A supervenience thesis is, in a broad sense, reductionist. But it is a stripped-down form or reductionism, unencumbered by dubious denials of existence, claims of ontological priority, or claims of translatibility. One might wish to say that in some sense the beauty of statues is nothing over and above the shape and size and colour that beholders appreciate, but without denying that there is such a thing as beauty, without claiming that beauty exists only in some less-than-fundamental way, and without under taking to paraphrase ascriptions of beauty in terms of shape etc. A supervenience thesis seems to capture what the cautious reductionist wishes to say.

Even if reductionists ought to be less cautious and aim for translation, still it is a good idea to attend to the question of supervenience. For if supervenience fails, then no scheme of translation can be correct and we needn't go on Chisholming away in search of one. If supervenience succeeds, on the other hand, then some correct scheme must exist; the remaining question is whether there exists a correct scheme that is less than infinitely complex. If beauty is supervenient on shape etc., the worst that can happen is that an ascription of beauty is equivalent to an uncountably infinite disjunction of maximally specific descriptions of shape etc., which descriptions might themselves involve infinite conjunctions.

Interesting supervenience theses usually involve the notion of qualitative duplication that we have just considered. Thus we may ask what does or doesn't supervene on the qualitative character of the entire world, throughout all of history. Suppose that two possible worlds are perfect qualitative duplicates — must they then also have exactly the same distributions of objective probability, the same laws of nature, the same counterfactuals and causal relations? Must their inhabitants have the same *de re* modal properties? If so, it makes sense to pursue such projects as a frequency analysis of probability, a regularity analysis of laws of nature, or a comparative similarity

---

[18]   For a general discussion of supervenience, see Jaegwon Kim, 'Supervenience and Nomological Incommensurables', *American Philosophical Quarterly* 15 (1978) pp. 149-156.

analysis of causal counterfactuals and *de re* modality. If not, such projects are doomed from the start, and we needn't look at the details of the attempts. But we cannot even raise these questions of supervenience unless we can speak of duplicate worlds. And to do that, I have suggested, we need natural properties.

(Note that if possible worlds obey a principle of identity of qualitative indiscernibles, then all these supervenience theses hold automatically. If no two worlds are duplicates, then *a fortiori* no two are duplicates that differ in their probabilities, laws, . . ., or anything else.)

We might also ask whether qualitative character supervenes on anything less. For instance, we might ask whether global qualitative character supervenes on local qualitative character. Say that two worlds are *local duplicates* iff they are divisible into corresponding small parts in such a way that (1) corresponding parts of the two worlds are duplicates, and (2) the correspondence preserves spatiotemporal relations. (The exact meaning depends, of course, on what we mean by 'small'.) If two worlds are local duplicates, then must they be duplicates *simpliciter*? Or could they differ in ways that do not prevent local duplication − *e.g.* in external relations, other than the spatiotemporal relations themselves, between separated things? Again, we must make sense of duplication − this time, both in the large and in the small − even to ask the question.[19]

* * *

Next, divergent worlds. I shall say that two possible worlds *diverge* iff they are not duplicates but they do have duplicate initial temporal segments. Thus our world and another might match perfectly up through the year 1945, and go their separate ways thereafter.

Note that we need no identity of times across worlds. Our world through our 1945 duplicates an initial segment of the other world; that otherworldly segment ends with a year that indeed resembles our 1945, but it is part of otherworldly time, not part of our time. Also, we need no separation of time and space that contravenes Relativity − we have initial temporal segments, of this or another world, if we have spatiotemporal regions bounded by spacelike surfaces that cut the world in two.

I distinguish *divergence* of worlds from *branching* of worlds. In branching, instead of duplicate segments, one and the same initial segment is allegedly shared as a common part by two overlapping worlds. Branching is problematic in ways that divergence is not. First, because an inhabitant of the shared segment cannot speak unequivocally of *the* world he lives in. What if he says there will be a sea fight tomorrow, meaning of course to speak of the future of his own world, and one of the two worlds he lives in has

[19] Such a thesis of supervenience of the global on the local resembles the 'holographic hypothesis' considered and rejected by Saul Kripke in 'Identity Through Time', presented at the 1979 conference of the American Philosophical Association, Eastern Division, and elsewhere.

a sea fight the next day and the other doesn't? Second, because overlap of worlds interferes with the most salient principle of demarcation for worlds, *viz.* that two possible individuals are part of the same world iff they are linked by some chain of external relations, *e.g.* of spatiotemporal relations. (I know of no other example.) Neither of these difficulties seems insuperable, but both are better avoided. That makes it reasonable to prefer a theory of nonoverlapping divergent worlds to a theory of branching worlds. Then we need to be able to speak of qualitative duplication of world-segments, which we can do in terms of shared natural properties.

Divergent (or branching) worlds are of use in defining Determinism. The usual definitions are not very satisfactory. If we say that every event has a cause, we overlook probabilistic causation under Indeterminism. If we speak of what could be predicted by a superhuman calculator with unlimited knowledge of history and the laws of nature, we overlook obstacles that might prevent prediction even under Determinism, or else we try to make nonvacuous sense of counterfactuals about what our predictor could do if he had some quite impossible combination of powers and limitations.

A better approach is as follows. First, a system of laws of nature is Deterministic iff no two divergent worlds both conform perfectly to the laws of that system. Second, a world is Deterministic iff its laws comprise a Deterministic system. Third, Determinism is the thesis that our world is Deterministic.[20]

(Alternative versions of Determinism can be defined in similar fashion. For instance, we could strengthen the first step by prohibiting convergence as well as divergence of law-abiding worlds. Or we could even require that no two law-abiding worlds have duplicate momentary slices without being duplicates throughout their histories. Or we could define a weaker sort of Determinism: we could call a world *fortuitously* Deterministic, even if its laws do not comprise a Deterministic system, iff no world both diverges from it and conforms to its laws. The laws and early history of such a world suffice to determine later history, but only because the situations in which the laws fall short of Determinism never arise. We might equivalently define fortuitous Determinism as follows: for any historical fact F and any initial segment S of the world, there are a true proposition H about the history of S and a true proposition L about the laws of nature, such that H and L together strictly imply F.[21] Does this definition bypass our need to speak of duplication of initial segments? Not so, for we must ask what it means to say that H

---

[20] This approach is due, in essence, to Richard Montague, 'Deterministic Theories', in *Decisions, Values and Groups,* II (Pergamon Press, 1962), and in his *Formal Philosophy* (Yale University Press, 1974). But Montague did not speak as I have done of duplication of initial segments of worlds in virtue of the sharing of certain élite properties. Instead, he used sameness of description in a certain vocabulary, which vocabulary was left as an unspecified parameter of his analysis. For he wrote as a logician obliged to remain neutral on questions of metaphysics.

[21] A closely related definition appears in Peter van Inwagen, 'The Incompatibility of Free Will and Determinism', *Philosophical Studies* 27 (1975) pp. 185-199.

except those recognised by physics. That is better, but I think still not right. Couldn't there be a natural property X (in the nature of the case, it is hard to name an example!) which is shared by the physical brains in worlds like ours and the immaterial spirits that inhabit other worlds? Or by thisworldly quarks and certain otherworldly particles that cannot exist under our physics? Physics could quite properly make no mention of a natural property of this sort. It is enough to recognise the special case applicable to our world, X-*cum*-physicality, brainhood or quarkhood as it might be. Then if by physical properties we mean those properties that are mentioned in the language of physics, a Materialist ought not to hold that all natural properties instantiated in our world are physical properties.

At this point, it ought to seem advisable to formulate Materialism as a supervenience thesis: no difference without physical difference. Or, contraposing: physical duplicates are duplicates *simpliciter. A fortiori,* no mental difference without physical difference; physical duplicates are mental duplicates. The thesis might best be taken as applying to whole possible worlds, in order to bypass such questions as whether mental life is to some extent extrinsic to the subject. So we have this first of several attempted formulations of Materialism:

> *M1.* Any two possible worlds that are exactly alike in all respects recognised by physics are qualitative duplicates.

But this will not do. In making Materialism into a thesis about how just any two worlds can and cannot differ, *M1* puts Materialism forward as a necessary truth. That is not what Materialists intend. Materialism is meant to be a contingent thesis, a merit of our world that not all other worlds share. Two worlds could indeed differ without differing physically, if a least one of them is a world where Materialism is false. For instance, our Materialistic world differs from a nonmaterialistic world that is physically just like ours but that also contains physically epiphenomenal spirits.

There is a noncontingent supervenience thesis nearby that might appeal to Materialists:

> *M2.* There is no difference, *a fortiori* no mental difference, without some nonmental difference. Any two worlds alike in all nonmental respects are duplicates, and in particular do not differ in respect of the mental lives of their inhabitants.

This seems to capture our thought that the mental is a pattern in a medium, obtaining in virtue of local features of the medium (neuron firings) and perhaps also very global features (laws of nature) that are too small or too big to be mental themselves. But *M2* is not Materialism. It is both less and more. Less, obviously, because it never says that the medium is physical. More, because it denies the very possibility of what I shall call *Panpsychistic* Materialism.

It is often noted that psychophysical identity is a two-way street: if all

mental properties are physical, then some physical properties are mental. But perhaps not just some but *all* physical properties might be mental as well; and indeed every property of anything might be at once physical and mental. Suppose there are indeed worlds where this is so. If so, presumably there are many such worlds, not all duplicates, differing *inter alia* in the mental lives of their inhabitants. But all differences between such worlds are mental (as well as physical), so none are nonmental. These worlds will be vacuously alike in all nonmental respects, for lack of any nonmental respects to differ in. Then *M2* fails. And not just at the troublemaking worlds; *M2* is noncontingent, so if it fails at any worlds, it fails at all — even decent Materialistic worlds like ours. Maybe Panpsychistic Materialism is indeed impossible — how do you square it with a broadly functional analysis of mind? — but a thesis that says so is more than just Materialism.

A third try. This much is at least true:

> *M3.* No two Materialistic worlds differ without differing physically; any two Materialistic worlds that are exactly alike physically are duplicates.

But *M3* is not a formulation of Materialism, for the distinction between Materialistic and other worlds appears within *M3*. All we learn is that the Materialistic worlds comprise a class within which there is no difference without physical difference. But there are many such classes. In fact any world, however spirit-ridden, belongs to such a class.

A fourth try. Perhaps we should confine our attention to nomologically possible worlds, thus:

> *M4.* Among worlds that conform to the actual laws of nature, no two differ without differing physically; any two such worlds that are exactly alike physically are duplicates.

But again we have something that is both less and more than Materialism. Less, because *M4* could hold at a world where Materialism is false but where spiritual phenomena are correlated with physical phenomena according to strict laws. More, because M4 fails to hold at a Materialistic, spirit-free world if the laws of that world do not preclude the existence of epiphenomenal spirits. Our world might be such a world, a world where spirits are absent but not outlawed.[23]

So far, a supervenience formulation of Materialism seems elusive. But I think we can succeed if we join the idea of supervenience with the idea that a nonmaterialistic world would have somthing extra, something that a Materialistic world lacks. It might have spirits; or it might have physical things that differ in nonphysical ways, for instance in what their experience is like.

---

[23] This objection against *M4* as a formulation of 'the ontological primacy of the microphysical' appears in Terence Horgan, 'Supervenience and Microphysics', *Pacific Philosophical Quarterly* 63 (1982) pp. 29-43.

In either case there are extra natural properties, properties instantiated in the nonmaterialistic world but nowhere to be found in the Materialistic world. Let us say that a property is *alien* to a world iff (1) it is not instantiated by any inhabitant of that world, and (2) it is not analysable as a conjunction of, or as a structural property constructed out of, natural properties all of which are instantiated by inhabitants of that world. (I need the second clause because I am following Armstrong, *mutatis mutandis,* in declining to rule out perfectly natural properties that are conjunctive or structurally complex. See *Universals,* II, pp. 30-42 and 67-71. It would be wrong to count as alien a complex property analysable in terms of nonalien constituents.) If our world is Materialistic, then it is safe to say that some of the natural properties instantiated in any nonmaterialistic world are properties alien to our world. Now we can proceed at last to formulate Materialism as a restricted and contingent supervenience thesis:

> *M5.* Among worlds where no natural properties alien to our world are instantiated, no two differ without differing physically; any two such worlds that are exactly alike physically are duplicates.[24]

<div align="center">* * *</div>

We took Materialism to uphold the comprehensiveness of 'something not too different from present-day physics, though presumably somewhat improved'. That was deliberately vague. Materialist metaphysicians want to side with physics, but not to take sides within physics. Within physics, more precise claims of completeness and correctness may be at issue. Physics (ignoring latter-day failures of nerve) is the science that aspires to comprehensiveness, and particular physical theories may be put forward as fulfilling that aspiration. If so, we must again ask what it means to claim comprehensiveness. And again, the answer may be given by a supervenience formulation: no difference without physical difference as conceived by such-and-such grand theory. But again it must be understood as a restricted and contingent supervenience thesis, applying only among worlds devoid of alien natural properties.

Thus the businesss of physics is not just to discover laws and causal explanations. In putting forward as comprehensive theories that recognise only a limited range of natural properties, physics proposes inventories of the natural properties instantiated in our world. Not complete inventories,

---

[24] This formulation resembles one proposed by Horgan, *op. cit.* The principal difference is as follows. Horgan would count as alien (my term, not his) any property cited in the fundamental laws of otherworldly microphysics that is not also explicitly cited in the fundamental laws of this worldly microphysics. Whether the property is instantiated in either world doesn't enter into it. But must an alien property figure in laws of otherworldly *physics?* Must it figure in any otherworldly laws at all? It seems that a Materialistic world might differ without differing physically from a world where there are properties alien in my sense but not in Horgan's — perhaps a world where laws are in short supply.

perhaps. But complete enough to account for all the duplications and differences that could arise in the absence of alien natural properties. Of course, the discovery of natural properties is inseparable from the discovery of laws. For an excellent reason to think that some hitherto unsuspected natural properties are instantiated — properties deserving of recognition by physics, the quark colours as they might be — is that without them, no satisfactory system of laws can be found.

This is reminiscent of the distinctive *a posteriori,* scientific character of Armstrong's Realism (*Universals,* I, pp. 8-9, and *passim*). But in the setting of an ontology of *possibilia,* the distinction between discovering what universals or natural properties there actually are and discovering which ones are actually instantiated fades away. And the latter question is *a posteriori* on any theory. What remains, and remains important, is that physics discovers properties. And not just any properties — natural properties. The discovery is, for instance, that neutrinos are not all alike. That is not the discovery that different ones have different properties in my sense, belong to different classes. We knew that much *a priori.* Rather, it is the surprising discovery that some *natural* property differentiates some neutrinos from others. That discovery has in fact been made; I should like to read an account of it by some philosopher who is not prepared to adopt a discriminatory attitude toward properties and who thinks that all things are equally similar and dissimilar to one another.

Laws and Causation. The observation that physics discovers natural properties in the course of discovering laws may serve to introduce our next topic: the analysis of what it is to be a law of nature. I agree with Armstrong that we need universals, or at least natural properties, in explaining what lawhood is, though I disagree with his account of how this is so.

Armstrong's theory, in its simplest form,[25] holds that what makes certain regularities lawful are second-order states of affairs $N(F,G)$ in which the two ordinary, first-order universals F and G are related by a certain dyadic second-order universal N. It is a contingent matter which universals are thus related by the lawmaker N. But it is necessary — and necessary *simpliciter,* not just nomologically necessary — that if $N(F,G)$ obtains, then F and G are constantly conjoined. There is a necessary connection between the second-order state of affairs $N(F,G)$ and the first-order lawful regularity $\forall x(Fx \supset Gx)$; and likewise between the conjunctive state of affairs $N(F,G)$ & Fa and its necessary consequence Ga.

A parallel theory could be set up with natural properties in place of Armstrong's first- and second-order universals. It would have many of the attractive features that Armstrong claims on behalf of his theory, but at least

[25] *Universals,* II, pp. 148-157. A more developed form of the theory appears in D. M. Armstrong, *What Is a Law of Nature?* (Cambridge University Press, 1983). Similar theories have been proposed in Fred I. Dretske, 'Laws of Nature', *Philosophy of Science* 44 (1977) pp. 248-268, and in Michael Tooley, 'The Nature of Laws', *Canadian Journal of Philosophy* 4 (1977) pp. 667-698.

one merit would be lost. For Armstrong, the lawful necessitation of *Ga* by *Fa* is a purely local matter: it involves only *a,* the universals *F* and *G* that are present in *a,* and the second-order lawmaking universal that is present in turn in (or between) these two universals. If we replace the universals by properties, however natural, that locality is lost. For properties are classes with their membership spread around the worlds, and are not wholly present in *a*. But I do not think this a conclusive objection, for our intuitions of locality often seem to lead us astray. The selective regularity theory I shall shortly advocate also sacrifices locality, as does any regularity theory of law.

What leads me (with some regret) to reject Armstrong's theory, whether with universals or with natural properties, is that I find its necessary connections unintelligible. Whatever *N* may be, I cannot see how it could be absolutely impossible to have *N(F,G)* and *Fa* without *Ga*. (Unless *N* just *is* constant conjunction, or constant conjunction plus something else, in which case Armstrong's theory turns into a form of the regularity theory he rejects.) The mystery is somewhat hidden by Armstrong's terminology. He uses 'necessitates' as a name for the lawmaking universal *N*; and who would be surprised to hear that if *F* 'necessitates' *G* and *a* has *F*, then *a* must have *G*? But I say that *N* deserves the name of 'necessitation' only if, somehow, it really can enter into the requisite necessary connections. It can't enter into them just by bearing a name, any more than one can have mighty biceps just by being called 'Armstrong'.

I am tempted to complain in Humean fashion of alleged necessary connections between distinct existences, especially when first-order states of affairs in the past supposedly join with second-order states of affairs to necessitate first-order states of affairs in the future. That complaint is not clearly right: the sharing of universals detracts from the distinctness of the necessitating and the necessitated states of affairs. But I am not appeased. I conclude that necessary connections can be unintelligible even when they are supposed to obtain between existences that are not clearly and wholly distinct.[26]

Thus I do not endorse Armstrong's way of building universals, or alternatively natural properties, into the analysis of lawhood. Instead I favour a regularity analysis. But I need natural properties even so.

Certainly not just any regularity is a law of nature. Some are accidental.

---

[26] Armstrong's more developed theory in *What Is a Law of Nature?* complicates the picture in two ways. First, the second order state of affairs *N(F,G)* is itself taken to be a universal, and its presence in its instances detracts yet further from the distinctness of the necessitating and the necessitated states of affairs. Second, all laws are defeasible. It is possible after all to have *N(F,G)* and *Fa* without *Ga*, namely if we also have *N(E&F,H)* and *Ea*, where *H* and *G* are incompatible. (The law that *F*'s are *G*'s might be *contingently* indefeasible, if no such defeating state of affairs *N(E&F,H)* obtains; but no law has its indefeasibility built in essentially.) It remains true that there are alleged necessary connections that I find unintelligible, but they are more complicated than before. To necessitate a state of affairs, we need not only the first- and second-order states of affairs originally considered, but also a negative existential to the effect that there are no further states of affairs of the sort that could act as defeaters.

So an adequate regularity analysis must be selective. Also, an adequate analysis must be collective. It must treat regularities not one at a time, but rather as candidates to enter into integrated systems. For a given regularity might hold either as a law or accidentally, depending on whether other regularities obtain that can fit together with it in a suitable system. (Thus I reject the idea that lawhood consists of 'lawlikeness' plus truth.) Following Mill and Ramsey,[27] I take a suitable system to be one that has the virtues we aspire to in our own theory-building, and that has them to the greatest extent possible given the way the world is. It must be entirely true; it must be closed under strict implication; it must be as simple in axiomatisation as it can be without sacrificing too much information content; and it must have as much information content as it can have without sacrificing too much simplicity. A law is any regularity that earns inclusion in the ideal system. (Or, in case of ties, in every ideal system.) The ideal system need not consist entirely of regularities; particular facts may gain entry if they contribute enough to collective simplicity and strength. (For instance, certain particular facts about the Big Bang might be strong candidates.) But only the regularities of the system are to count as laws.

We face an obvious problem. Different ways to express the same content, using different vocabulary, will differ in simplicity. The problem can be put in two ways, depending on whether we take our systems as consisting of propositions (classes of worlds) or as consisting of interpreted sentences. In the first case, the problem is that a single system has different degrees of simplicity relative to different linguistic formulations. In the second case, the problem is that equivalent systems, strictly implying the very same regularites, may differ in their simplicity. In fact, the content of any system whatever may be formulated very simply indeed. Given system S, let F be a predicate that applies to all and only things at worlds where S holds. Take F as primitive, and axiomatise S (or an equivalent thereof) by the single axiom $\forall x F x$. If utter simplicity is so easily attained, the ideal theory may as well be as strong as possible. Simplicity and strength needn't be traded off. Then the ideal theory will include (its simple axiom will strictly imply) all truths, and *a fortiori* all regularities. Then, after all, every regularity will be a law. That must be wrong.

The remedy, of course, is not to tolerate such a perverse choice of primitive vocabulary. We should ask how candidate systems compare in simplicity when each is formulated in the simplest eligible way; or, if we count different formulations as different systems, we should dismiss the ineligible ones from candidacy. An appropriate standard of eligibility not far to seek: let the

[27] John Stuart Mill, *A System of Logic* (Parker, 1843) Book III, Chapter IV, Section 1; F. P. Ramsey, 'Universals of Law and of Fact', in his *Foundations* (Routledge & Kegan Paul, 1978). Ramsey regarded this theory of law as superseded by the different theory in his 'General Propositions and Causality', also in *Foundations,* but I prefer his first thoughts to his second. I present a theory of lawhood along the lines of Ramsey's earlier theory in my *Counterfactuals* (Blackwell, 1973) pp. 73-75. A revision to that discussion is needed in the probabilistic case, which I here ignore.

primitive vocabulary that appears in the axioms refer only to perfectly natural properties.

Of course, it remains an unsolved and difficult problem to say what simplicity of a formulation is. But it is no longer the downright insoluble problem that it would be if there were nothing to choose between alternative primitive vocabularies.

(One might think also to replace strict implication by deducibility in some specified calculus. But this second remedy seems unnecessary given the first, and seems incapable of solving our problem by itself.)

If we adopt the remedy proposed, it will have the consequence that laws will tend to be regularities involving natural properties. Fundamental laws, those that the ideal system takes as axiomatic, must concern perfectly natural properties. Derived laws that follow fairly straightforwardly also will tend to concern fairly natural properties. Regularities concerning unnatural properties may indeed be strictly implied, and should count as derived laws if so. But they are apt to escape notice even if we someday possess a good approximation to the ideal system. For they will be hard to express in a language that has words mostly for not-too-unnatural properties, as any language must. (See the next section.) And they will be hard to derive, indeed they may not be finitely derivable at all, in our deductive calculi. Thus my account explains, as Armstrong's does in its very different way, why the scientific investigation of laws and of natural properties is a package deal; why physicists posit natural properties such as the quark colours in order to posit the laws in which those properties figure, so that laws and natural properties get discovered together.

*   *   *

If the analysis of lawhood requires natural properties, then so does the analysis of causation. It is fairly uncontroversial that causation involves laws. That is so according to both of the leading theories of causation: the deductive-nomological analysis, on which the laws are applied to the actual course of events with the cause and effect present; and the counterfactual analysis that I favour, on which the laws are applied to counterfactual situations with the cause hypothesised away. These counterfactual alternatives may need to break actual laws at the point where they diverge from actuality, but the analysis requires that they evolve thereafter in accordance with the actual laws.[28]

According to my counterfactual analysis, causation involves natural properties in a second way too. We need the kind of counterfactuals that avoid backtracking; else the analysis faces fatal counterexamples involving epiphenomenal side-effects or cases of causal preemption. As I have already noted, these counterfactuals are to be characterised in terms of divergent

---

[28]   See my 'Causation', *Journal of Philosophy* 70 (1973) pp. 556-567; reprinted in Ernest Sosa, ed., *Causation and Conditionals* (Oxford University Press, 1975).

worlds, hence in terms of duplicate initial world-segments, hence in terms of shared natural properties.

Causation involves natural properties in yet another way. (Small wonder that I came to appreciate natural properties after working on the analysis of causation!) Causation holds between events. Unless we distinguish genuine from spurious events, we will be left with too many putative causes. You put a lump of butter on a skillet, and the butter melts. What event causes this? There is one event that we can call a moving of molecules. It occurs in the region where the skillet is, just before the butter melts. This is an event such that, necessarily, it occurs in a spatiotemporal region only if that region contains rapidly moving molecules. Surely this event is a cause of the melting of the butter.

Heat is that phenomenon, whatever it may be, that manifests itself in certain familiar characteristic ways. Let us say: heat is that which occupies the heat-role. (It won't matter whether we take the definite description plain, as I prefer, or rigidified.) In fact, but contingently, it is molecular motion that occupies the heat-role. It might have been molecular nonmotion, or caloric fluid, or what you will. Now consider an alleged second event, one that we may call a having-the-occupant-of-the-heat-role. This second event occurs just when and where the first does, in the region where the hot skillet is. It occurs there in virtue of the two facts (1) that the skillet's molecules are moving rapidly, and (2) that the region in question is part of a world where molecular motion is what occupies the heat-role. But this second event differs from the first. The necessary conditions for its occurrence are different. Necessarily, it occurs in a region only if that region contains whatever phenomenon occupies the heat-role in the world of which that region is part. So in those worlds where caloric fluid occupies the heat-role and molecular motion does not, the first event occurs only in regions with molecular motion whereas the second occurs only in regions with caloric fluid.

Certainly the first event causes the melting of the butter, but shall we say that the second event does so as well? No; that seems to multiply causes beyond belief by playing a verbal trick. But if there really are two events here, I cannot see why the second has less of a claim than the first to be a cause of the melting of the butter. It is out of the question to say that the first and the second events are one and the same — then this one event would have different conditions of occurrence from itself. The best solution is to deny that the alleged second event is a genuine event at all. If it isn't, of course it can't do any causing.

Why is the first event genuine and the second spurious? Compare the properties involved: containing rapidly moving molecules versus containing whatever phenomenon occupies the heat-role. (I mean these as properties of the spatiotemporal region; other treatments of events would take instead the corresponding properties of the skillet, but my point would still apply.) The first is a fairly natural, intrinsic property. The second is highly disjunctive and extrinsic. For all sorts of different phenomena could occupy the heat-role; and whether the phenomenon going on in a region occupies the role

depends not only on what goes on in the region but also on what goes on elsewhere in the same world. Thus the distinction between more and less natural properties gives me the distinction between genuine and spurious events that I need in order to disown an overabundance of causes. If a property is too unnatural, it is inefficacious in the sense that it cannot figure in the conditions of occurrence of the events that cause things.[29]

The Content of Language and Thought. Hilary Putnam has given an argument which he regards as a refutation of a 'radically non-epistemic' view of truth, but which I regard rather as a *reductio* against Putnam's premises.[30] In particular, it refutes his assumption that '*we* interpret our languages or nothing does' ('Models and Reality', p. 482) so that any constraint on reference must be established by our own stipulation in language or thought. Gary Merrill has suggested that Putnam may be answered by appeal to a constraint that depends on an objective structure of properties and relations in the world.[31] I agree, and find here another point at which we need natural properties.

Putnam's argument, as I understand it, is as follows. First, suppose that the only contraint on interpretation of our language (or perhaps our language of thought) is given by a description theory of reference of a global and futuristic sort. An 'intended interpretation' is any interpretation that satisfies a certain body of theory: *viz.* the idealised descendant of our current total theory that would emerge at the end of inquiry, an ideal theory refined to perfection under the guidance of all needed observation and our best theoretical reasoning. If so, intended interpretations are surprisingly abundant. For *any* world can satisfy *any* theory (ideal or not), and can do so in countless very different ways, provided only that the world is not too small and the theory is consistent. Beyond that, it doesn't matter what the world is like or what the theory says. Hence we have radical indeterminacy of reference. And we have the coincidence that Putnam welcomes between satisfaction under all intended intrepretations and 'epistemic truth'. For the ideal theory is the whole of 'epistemic truth', the intended interpretations are just those interpretations of our language that satisfy the ideal theory, and

---

[29] See the discussion of impotence of dispositions in Elizabeth W. Prior, Robert Pargetter, and Frank Jackson, 'Three Theses About Dispositions', *American Philosophical Quarterly* 19 (1982) pp. 251-257. If a disposition is not identified with its actual basis, there is a threat of multiplication of putative causes similar to that in my example. We would not wish to say that the breaking of a struck glass is caused both by its fragility and by the frozen-in stresses that are the basis thereof; and if forced to choose, we should choose the latter. I suggest that the fragility is inefficacious because it is too unnatural a property, too disjunctive and extrinsic, to figure in the conditions of occurence of any event.

[30] Hilary Putnam, 'Realism and Reason', in his *Meaning and the Moral Sciences* (Routledge & Kegan Paul, 1978), and 'Models and Reality', *Journal of Symbolic Logic* 45 (1980) pp. 464-482. The reader is warned that the argument as I present it may not be quite as Putnam intended it to be. For I have made free in reading between the lines and in restating the argument in my own way.

[31] G. H. Merrill, 'The Model-Theoretic Argument Against Realism', *Philosophy of Science* 47 (1980) pp. 69-81.

(unless the world is too small or ideal theory is inconsistent) there are some such interpretations.

I take this to refute the supposition that there are no further constraints on reference. But Putnam asks: how *could* there be a further constraint? How could we ever establish it? By stipulation, by saying or thinking something. But whatever we say or think will be in language (or language of thought) that suffers from radical indeterminacy of interpretation. For the saving constraint will not be there until we succeed in establishing it. So the attempted stipulation must fail. The most we can do is to contribute a new chapter to current and ideal theory, a chapter consisting of whatever we said or thought in our stipulation. And this new theory goes the way of all theory. So we cannot establish a further constraint; and 'we interpret our language or nothing does'; so there cannot be any further constraint. We cannot lift ourselves by our bootstraps, so we must still be on the ground.

Indeed we cannot lift ourselves by our bootstraps, but we are off the ground, so there must be another way to fly. Our language does have a fairly determinate interpretation (a Moorean fact!) so there must be some constraint not created *ex nihilo* by our stipulation.

What can it be? Many philosophers would suggest that it is some sort of causal constraint. If so my case is made, given my arguments in the previous section: we need natural properties to explain determinacy of interpretation. But I doubt that it really is a causal constraint, for I am inclined to think that the causal aspect of reference *is* established by what we say and think. Thus: I think of a thing as that which I am causally acquainted with in such-and-such way, perhaps perceptually or perhaps through a channel of acquaintance that involves the naming of the thing and my picking up of the name. I refer to that thing in my thought, and derivatively in language, because it is the thing that fits this causal and egocentric description extracted from my theory of the world and of my place in the world.[32]

I would instead propose that the saving constraint concerns the referent – not the referrer, and not the causal channels between the two. It takes two to make a reference, and we will not find the constraint if we look for it always on the wrong side of the relationship. Reference consists in part of what we do in language or thought when we refer, but in part it consists in eligibility of the referent. And this eligibility to be referred to is a matter of natural properties.

That is the suggestion Merrill offers. (He offers it not as his own view, but as what opponents of Putnam ought to say; and I gratefully accept the offer.) In the simplest case, suppose that the interpretation of the logical vocabulary somehow takes care of itself, to reveal a standard first-order language whose nonlogical vocabulary consists entirely of predicates. The parts of the world comprise a domain; and sets, sets of pairs, . . ., from this domain are potential extensions for the predicates. Now suppose we have

---

[32] See Stephen Schiffer, 'The Basis of Reference', *Erkenntnis* 13 (1978) pp. 171-206.

an all-or-nothing division of properties into natural and unnatural. Say that a set from the domain is *eligible* to be the extension of a one-place predicate iff its members are just those things in the domain that share some natural property; and likewise for many-place predicates and natural relations. An *eligible interpretation* is one that assigns none but eligible extensions to the predicates. A so-called 'intended' interpretation is an eligible interpretation that satisfies the ideal theory. (But the name is misleading: it is not to be said that our intentions establish the constraint requiring eligibility. That way lies the futile bootstrap-tugging that we must avoid.) Then if the natural properties are sparse, there is no reason to expect any overabundance of intended interpretations. There may even be none. Even ideal theory runs the risk of beings unsatisfiable, save in 'unintended' ways. Because satisfaction is not guaranteed, we accomplish something if we manage to achieve it by making a good fit between theory and the world. All this is as it should be.

The proposal calls for refinement. First, we need to provide for richer forms of language. In this we can be guided by familiar translations, for instance between modal language with higher-order quantification and first-order language that explicitly mentions *possibilia* and classes built up from them. Second, it will not do to take naturalness of properties as all-or-nothing. Here, above all, we need to make naturalness − and hence eligibility − a comparative matter, or a matter of degree. There are salient sharp lines, but not in the right places. There is the line between the perfectly natural properties and all the rest, but surely we have predicates for much-less-than-perfectly natural properties. There is the line between properties that are and that are not finitely analysable in terms of perfectly natural properties, but that lets in enough highly unnatural properties that it threatens not to solve our problem. We need gradations; and we need some give and take between the eligibility of referents and the other factors that make for 'intendedness', notably satisfaction of appropriate bits of theory. (Ideal theory, if we keep as much of Putnam's story as we can.) Grueness is not an absolutely ineligible referent (as witness my reference to it just now) but an interpretation that assigns it is to that extent inferior to one that assigns blueness instead. *Ceteris paribus,* the latter is the 'intended' one, just because it does better on eligibility.

Naturalness of properties makes for differences of eligibility not only among the properties themselves, but also among things. Compare Bruce with the cat-shaped chunk of miscellaneous and ever-changing matter that follows him around, always a few steps behind. The former is a highly eligible referent, the latter is not. (I haven't succeeded in referring to it, for I didn't say just which such chunk 'it' was to be.) That is because Bruce, unlike the cat-shaped chunk, has a boundary well demarcated by differences in highly natural properties. Where Bruce ends, there the density of matter, the relative abundance of the chemical elements, . . . abruptly change. Not so for the chunk. Bruce is also much more of a locus of causal chains than is the chunk; this too traces back to natural properties, by the considerations of the previous section. Thus naturalness of properties sets up distinctions among things. The reverse happens also. Once we are away from the perfectly natural

properties, one thing that makes for naturalness of a property is that it is a property belonging exclusively to well-demarcated things.

* * *

You might well protest that Putnam's problem is misconceived, wherefore no need has been demonstrated for resources to solve it. Putnam seems to conceive of language entirely as a repository of theory, and not at all as a practice of social interaction. We have the language of the encyclopedia, but where is the language of the pub? Where are the communicative intentions and the mutual expectations that seem to have so much to do with what we mean? In fact, where is thought? It seems to enter the picture, if at all, only as the special case where the language to be interpreted is hard-wired, unspoken, hidden, and all too conjectural.

I think the point is well taken, but I think it doesn't matter. If the problem of inentionality is rightly posed there will still be a threat of radical indeterminacy, there will still be a need for saving constraints, there will still be a remedy analogous to Merrill's suggested answer to Putnam, and there will still be a need for natural properties.

Set language aside and consider instead the interpretation of thought. (Afterward we can hope to interpret the subject's language in terms of his beliefs and desires regarding verbal communication with others.) The subject is in various states, and could be in various others, that are causally related to each other, to the subject's behaviour, and to the nearby environment that stimulates his senses. These states fit into a functional organisation, they occupy certain causal roles. (Most likely they are states of the brain. Maybe they involve something that is language-like but hard-wired, maybe not. But the nature of the states is beside the point.) The states have their functional roles in the subject as he now is, and in the subject as he is at other times and as he might have been under other circumstances, and even in other creatures of the same kind as the subject. Given the functional roles of the states, the problem is to assign them content. Propositional content, some would say; but I would agree only if the propositions can be taken as egocentric ones, and I think an 'egocentric proposition' is simply a property. States indexed by content can be identified as a belief that this, a desire for that, a perceptual experience of seeming to confront so-and-so, an intention to do such-and-such. (But not all ordinary ascriptions of attitudes merely specify the content of the subject's states. Fred and Ted might be alike in the functional roles of their states, and hence have states with the same content in the narrowly psychological sense that is my present concern, and hence believe alike *e.g.* by each believing himself to have heard of a pretty town named 'Castlemaine'. Yet they might be acquainted via that name with different towns, at opposite ends of the earth, so that Fred and not Ted believes that Castlemaine, Victoria, is pretty.) The problem of assigning content to functionally characterised states is to be solved by means of

constraining principles. Foremost among these are principles of fit. If a state is to be interpreted as an intention to raise one's hand, it had better typically cause the hand to go up. If a state (or complex of states) is to be interpreted as as a system of beliefs and desires — or better, degrees of belief and desire — according to which raising one's hand would be a good means to one's ends, and if another state is to be interpreted as an intention to raise one's hand, then the former had better typically cause the latter. Likewise on the input side. A state typically caused by round things before the eyes is a good candidate for interpretation as the visual experience of confronting something round; and its typical impact on the states interpreted as systems of belief ought to be interpreted as the exogenous addition of a belief that one is confronting something round, with whatever adjustment that addition calls for.

So far, so good. But it seems clear that preposterous and perverse misinterpretations could nevertheless cohere, could manage to fit the functional roles of the states because misassignment of content at one point compensates for misassignment at another. Let us see just how this could happen, at least under an oversimplified picture of interpretation as follows. An interpretation is given by a pair of functions C and V. C is a probability distribution over the worlds, regarded as encapsulating the subject's dispositions to form beliefs under the impact of sensory evidence: if a stream of evidence specified by proposition E would put the subject into a total state S — for short, if E *yields* S — we interpret S to consist in part of the belief system given by the probability distribution $C(-/E)$ that comes from C by conditionalising on E. V is a function from worlds to numerical desirability scores, regarded as encapsulating the subject's basic values: if E yields S, we interpret S to consist in part of the system of desires given by the $C(-/E)$-expectations of V. Say that C and V *rationalise* behaviour B after evidence E iff the system of desires given by the $C(-/E)$-expectations of V ranks B at least as high as any alternative behaviour. Say that C and V *fit* iff, for any evidence-specifying E, E yields a state that would cause behaviour rationalised by C and V after E. That is our only constraining principle of fit. (Where did the others go? — We built them into the definitions whereby C and V encapsulate an assignment of content to various states.) Then any two interpretations that always rationalise the same behaviour after the same evidence must fit equally well. Call two worlds *equivalent* iff they are alike in respect of the subject's evidence and behaviour, and note that any decent world is equivalent *inter alia* to horrendously counterinductive worlds and to worlds where everything unobserved by the subject is horrendously nasty. Fit depends on the total of C for each equivalence class, and on the C-expectation of V within each class, but that is all. Within a class, it makes no difference which world gets which pair of values of C and V. We can interchange equivalent worlds *ad lib* and preserve fit. So, given any fitting and reasonable interpretation, we can transform it into an equally fitting perverse interpretation by swapping equivalent worlds around so as to enhance

the probabilities of counterinductive worlds, or the desirabilities of nasty worlds, or both. *Quod erat demonstrandum.*

(My simplifications were dire: I left out the egocentricity of belief and desire and evidence, the causal aspect of rationalised behaviour, the role of intentions, change of basic values, limitations of logical competence, ... But I doubt that these omissions matter to my conclusion. I conjecture that if they were remedied, we could still transform reasonable interpretations into perverse ones in a way that preserves fit.)

If we rely on principles of fit to do the whole job, we can expect radical indeterminacy of interpretation. We need further constraints, of the sort called principles of (sophisticated) charity, or of 'humanity'.[33] Such principles call for interpretations according to which the subject has attitudes that we would deem reasonable for one who has lived the life that he has lived. (Unlike principles of crude charity, they call for imputations of error if he has lived under deceptive conditions.) These principles select among conflicting interpretations that equally well conform to the principles of fit. They impose *a priori* — albeit defeasible — presumptions about what sorts of things are apt to be believed and desired; or rather, about what dispositions to develop beliefs and desires, what inductive biases and basic values, someone may rightly be interpreted to have.

It is here that we need natural properties. The principles of charity will impute a bias toward believing that things are green rather than grue, toward having a basic desire for long life rather than for long-life-unless-one-was-born-on-Monday-and-in-that-case-life-for-an-even-number-of-weeks.   In short, they will impute eligible content, where ineligibility consists in severe unnaturalness of the properties the subject supposedly believes or desires or intends himself to have. They will impute other things as well, but it is the imputed eligibility that matters to us at present.

Thus the threat of radical indeterminacy in the assignment of content to thought is fended off. The saving constraint concerns the content — not the thinker, and not any channels between the two. It takes two to index states with content, and we will not find the constraint if we look for it always on the wrong side of the relationship. Believing this or desiring that consists in part in the functional roles of the states whereby we believe or desire, but in part it consists in the eligibility of the content. And this eligibility to be thought is a matter, in part, of natural properties.

Consider the puzzle whereby Kripke illustrates Wittgenstein's paradox that 'no course of action could be determined by a rule, because every course of action can be made out to accord with the rule'.[34] A well-educated person working arithmetic problems intends to perform addition when he sees the '+' sign. He does not intend to perform quaddition, which is just like addition

---

[33] See my 'Radical Interpretation', *Synthese* 23 (1974) pp. 331-344; and Richard E. Grandy, 'Reference, Meaning and Belief', *Journal of Philosophy* 70 (1973) pp. 439-452.

[34] See Saul A. Kripke, 'Wittgenstein on Rules and Private Language: An Elementary Exposition', in Irving Block, ed., *Perspectives on Wittgenstein* (Blackwell, 1981).

for small numbers but which yields the answer 5 if any of the numbers to be quadded exceeds a certain bound. Wherefore does he intend to add and not to quadd? Whatever he says and whatever is written in his brain can be perversely (mis)interpreted as instructing him to quadd. And it is not enough to say that his brain state is the causal basis of a disposition to add. Perhaps it isn't. Perhaps if a test case arose he would abandon his intention, he would neither add nor quadd but instead would put his homework aside and complain that the problems are too hard.

The naive solution is that adding means going on in the same way as before when the numbers get big, whereas quadding means doing something different; there is nothing present in the subject that constitutes an intention to do different things in different cases; therefore he intends addition, not quaddition. We should not scoff at this naive response. It is the correct solution to the puzzle. But we must pay to regain our naiveté. Our theory of properties must have adequate resources to somehow ratify the judgement that instances of adding are all alike in a way that instances of quadding are not. The property of adding is not perfectly natural, of course, not on a par with unit charge or sphericality. And the property of quadding is not perfectly unnatural. But quadding is worse by a disjunction. So quaddition is to that extent less of a way to go on doing the same, and therefore it is to that extent less of an eligible thing to intend to do.

It's not that you couldn't possibly intend to quadd. You could. Suppose that today there is as much basis as there ever is to interpret you as intending to add and as meaning addition by your word 'addition' and quaddition by 'quaddition'; and tomorrow you say to yourself in so many words that it would be fun to tease the philosophers by taking up quadditon henceforth, and you make up your mind to do it. But you have to go out of your way. Adding and quadding aren't on a par. To intend to add, you need only have states that would fit either interpretation and leave it to charity to decree that you have the more eligible intention. To intend to quadd, you must say or think something that creates difficulties of fit for the more eligible intention and thereby defeats the presumption in its favour. You must do something that, taking principles of fit and presumptions of eligibility and other principles of charity together, tilts the balance in favour of an interpretation on which you intend to quadd. How ironic that we were worried to find nothing positive to settle the matter in favour of addition! For the lack of anything positive that points either way just *is* what it takes to favour addition. Quaddition, being less natural and eligible, needs something positive in its favour. Addition can win by default.

What is the status of the principles that constrain interpretation, in particular the charitable presumption in favour of eligible content? We must shun several misunderstandings. It is not to be said (1) that as a contingent psychological fact, the contents of our states turn out to be fairly eligible, we mostly believe and desire ourselves to have not-too-unnatural properties. Still less should it be said (2) that we should daringly presuppose this in our interpreting of one another, even if we haven't a shred of evidence for it.

Nor should it be said (3) that as a contingent psychological fact we turn out to have states whose content involves some properties rather than others, and that is what makes it so that the former properties are more natural. (This would be a psychologistic theory of naturalness.) The error is the same in all three cases. It is supposed, wrongly as I think, that the problem of interpretation can be solved without bringing to it the distinction between natural and unnatural properties; so that the natural properties might or might not turn out to be the ones featured in the content of thought according to the correct solution, or so that they can afterward be defined as the ones that are so featured. I think this is overoptimistic. We have no notion how to solve the problem of interpretation while regarding all properties as equally eligible to feature in content. For that would be to solve it without enough contraints. Only if we have an independent, objective distinction among properties, and we impose the presumption in favour of eligible content *a priori* as a constitutive constraint, does the problem of interpretation have any solution at all. If so, then any correct solution must automatically respect the presumption. There's no contingent fact of psychology here to be believed, either on evidence or daringly.

Compare our selective and collective theory of lawhood: lawhood of a regularity just consists in its fitting into an ideally high-scoring system, so it's inevitable that laws turn out to have what it takes to make for high scores. Likewise, I have suggested, contenthood just consists in getting assigned by a high-scoring interpretation, so it's inevitable that contents tend to have what it takes to make for high scores. And in both cases, I've suggested that part of what it takes is naturalness of the properties involved. The reason natural properties feature in the contents of our attitudes is that naturalness is part of what it is to feature therein. It's not that we're built to take a special interest in natural properties, or that we confer naturalness on properties when we happen to take an interest in them.[35]

Princeton University