

9/12/2006

Naturalism Week 2 (These are *very rough* notes):

Part I: Plan (in 4 parts):

1. General Set-up:

- a) Naturalism as seeking to demonstrate or articulate a relation between a naturalistic base vocabulary and some target vocabulary or vocabularies.

Hence:

- i. Problems about defining the *base* vocabulary: physicalism or broadly scientific naturalism or extended-nature naturalism; how widely to draw the boundaries around 'science'; the Hempel-Crane&Mellor problem (which I think of as *not* impugning the credentials of the problem).
- ii. Issue of what *target* vocabulary one considers: normative (if so, especially moral?), psychological (if so, sentience or sapience, i.e. consciousness or intentionality?).
- iii. Issue of how to construe the *relation*: reduction (in the Carnap-Nagel sense), partial reduction, type-type identities (with or without derivability of laws), token-token identities (presumably, without derivability of laws), supervenience of various kinds. This is our topic this week.

- b) A way to think about the most general issue of physicalism (or other naturalism, whether broadly scientific or extended, is in terms of two questions:

- i. What sort of privilege does physics (or other favored naturalistic base vocabulary) have *within its own domain*? Possible answers: *causal* completeness, *explanatory* completeness. Lewis's formulation:

(From Menzies & Price on the Canberra Plan):

As for the second premise, Lewis characterises it as "the explanatory adequacy of physics". (1966, p. 23) It is the principle that "there is some unified body of scientific theory, of the sort we now accept, which together provide a true and exhaustive account of all physical phenomena." (1966, p. 23) As Lewis goes on to explain, to assume this principle is not to assume physicalism itself: "My second premise does not rule out the existence of nonphysical phenomena; it is not an ontological thesis in its own right. It only denies that we need ever explain physical phenomena by nonphysical ones." (pp. 23–24)

- ii. What sort of privilege (e.g. ontological, ideological) should we accord physics (or other favored naturalistic base vocabulary) over *other* discursive domains *in virtue of* the sort of privilege (i) it has in its *own* domain?
- iii. These two questions ((i) and (ii)) are surprisingly intimately linked. Causal or explanatory completeness says something like [look at various formulations, including Lewis's]: if you have a whole theory, only some of the terms of which are in

the vocabulary of physics, one can give wholly adequate explanations of everything describable just in the language of physics in that very language. That is, one need not appeal to anything non-physical to explain anything physical. This is already a claim to which the Craig Theorem is directly relevant, when it is considered as generally as I want to do. So *maybe* the Craig theorem is relevant to (a), and the Beth theorem to (b).

- iv. Q: Why should we think that strong global supervenience, let us say, of the intentional and the normative, is *true*?
- v. A possible answer is to be found in the causal (as we will see, it would be better to say *explanatory*) completeness of physics.
- vi. Notice that if we fill in this reason by saying that every event-object-particular also has a physical specification (a very strong and potentially contestable claim), and so is explicable under that specification, we have climbed back up from supervenience to token-token identity, of the Sellars-Davidson sort.
- vii. Notice that, as Russell had already pointed out in the 20's, cause is not a concept that appears in the official mathematized theories of mature fundamental physics.
- viii. And cause is a concept that would have to be substantially stretched from its informal Newtonian and pre-Newtonian senses to apply to *statistical* relations (as in statistical thermodynamics and QM). Should one be comfortable saying that the probability of A being  $p_1$  *causes* the probability of B being  $p_2$ ?
- ix. If we do allow "statistical causation" (where the idea of an intervening *mechanism* may have to be still further stretched, or even abandoned), we have to give up the idea that if A causes B, then if A were to happen, B would happen. For the inferences that are supported are not these strong counterfactuals, but only statistical-probabilistic inferences. And the possibility of *overdetermination* means that we must in any case give up the idea that if A causes B, then if A had *not* happened, then B *would not* have happened.
- x. It seems that what we are dealing with is the concept of a special kind of *explanation*, of which strictly *causal* explanation is perhaps a paradigmatic species, but not the *only* one.

c) How we got where we are: Reductionism done in by two sorts of arguments:

- i. Many levels argument: even *physics* has multiple, irreducible levels. Putnam on round pegs in square holes: "Many levels": Even within the triad of *general* natural sciences, physics,

chemistry, biology, definitional-descriptive and explanatory reductionism do not seem to be true. Here is Putnam's example as characterized by Block ("Anti-Reductionism Slaps Back", in *Philosophical Perspectives* [<http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/AntiReductionism.html>]):

Another influential analogy was Putnam's (1975) explanation of why a solid rigid round peg 1 inch in diameter won't fit through a square hole in a solid rigid board with a 1 inch diagonal. We can contrast the "upper level" explanation in terms of solidity, rigidity and geometry with the "lower level" account in terms of the specific elementary particle constitutions of a specific metal peg and wooden board. The upper level account is more general in that it applies to any solid rigid peg and board with that geometry, including materials that are composed of glass (a supercooled liquid) instead of the lattice structure of metals or the organic cell structure of wood. But the lower level account is more powerful in that it explains the specific cases of solidity and rigidity themselves. Further, it is more general because it explains details of the interaction between the peg and the board, including cases where the peg crumbles or the board breaks or tears.

- ii. Multiple realizability arguments: which we'll consider along with Kim and Fodor. Here is Block again:

Fodor and Putnam initiated the anti-reductionist consensus thirty years ago by noting the analogy between computational states and mental states (Fodor, 1965, 1974; Putnam, 1965, 1967). Any computational property can be "realized" or "implemented" in a variety of ways (electronic, mechanical, hydraulic), so it would be a mistake to identify any computational property with, say, an electronic property, since the same computational property can be implemented *without* the electronic property, for example mechanically. If thought is computational or functional, then for the same reason it would be a mistake to identify thought with any neural state; for thought can be implemented non-neurally, e.g. electronically. It would be wrong to identify thinking with a brain state if a device without a brain could think.

- d) It can be hard to see what the difference and exact relation between these points is. But our discussion of Kim will show this, I think.
- e) Praise the Haugeland, Kim, and Fodor papers for their philosophical style:
  - i. their sensible,
  - ii. sequential construction,
  - iii. with the issues, considerations, and previous crucial literature crisply and fairly laid out,
  - iv. their own contributions presented and assessed,
  - v. all in simple, straightforward sentences.
- f) Also—and maybe Kim is the best example of this—some issues of definition are noticed but put aside, since a rough-and-ready formulation is good enough to be getting on with. This is one of the myriad of dimensions along which good judgment must be made in order to keep from spinning one's wheels: where to insist on precision or nothing, and where not.

- i. Example: I judge that of the two sorts of considerations about the naturalistic *base* vocabulary (as opposed to worries about the nature of the *relations* between base and target vocabularies: e.g. reduction vs. supervenience) that the trouble physicalists have (Hempel and Crane&Mellor) saying *which* physics it is that they mean, temporally or ideally, while worth keeping in mind, is *not* the sort that ought to make us conclude that physicalism is simply not worth worrying about, while the issue of demarcation of the ‘science’ that forms the base language in terms of its *extent*, i.e. to what extent *special natural sciences* and *social sciences* should be included *is* worth taking very seriously.
- ii. Another example of a distinct but related phenomenon: I judge that the distinction between *modally weak* and *strong* supervenience is *not* important. (Kim’s notion of modally weak supervenience is a straw-man.) But the distinction between *global* supervenience (Haugeland’s “weak supervenience”) and *local* (or sometimes ‘*regional*’) supervenience is of the utmost importance.

g) t

2. Haugeland: Haugeland offers two big arguments:

- a) Here must discuss global vs. local or ‘regional’ supervenience, and mention Kim’s modally strong vs. weak.
- b) First argument: One vocabulary or range of facts (statable in some vocabulary) can *globally* (He says ‘weakly’, but this collides with Kim’s use of what we can call ‘modally weak’ supervenience. I agree with Horgan that this is a strawman.) *supervene* on another without there being token-token identities between them. This is an argument that shows *consistency* of supervenience & not token-token identity by providing a *model* that satisfies the one and not the other. This is facts-without-objects physicalism (say). Contrast Davidson’s objects-without-facts token-token identity theory.
- c) Second argument: An argument against the plausibility of token-token identity theories, in the psychology(intentionality)-physics: once again, an example of a case in *physics* where token-token identity of *macroscopic* and *microscopic* events does *not* hold. If it doesn’t hold even there, why should we think it does for psychology? Note: this is also a many-levels argument.

3. Kim (and Fodor):

- a) Wonderfully clear summary of the rise of the consensus against reductionism based on the possibility of multiply-realized *functional* concepts-properties.
- b) Argues that multiply realized functional properties, if thought of in terms of their realizations, would be seriously *disjunctive*.

- c) Q: What is wrong with disjunctive concepts? Notice that you can't tell just *syntactically*: jade is a disjunctive concept, in the sense in which they are objectionable, and African emerald or non-African emerald is *not*. (But what about African elephant or non-African elephant? Here both disjuncts are genuine kinds, and so is the disjunction. Fodor makes this point.)
  - d) A: They don't play the right role in *laws*, hence are not *inductively projectible*.
  - e) Kim claims that *if* psychological properties *are* functional and so multiply realized, then there cannot be a science of psychology.
  - f) *Unless* it is a disjunctive science of the different sorts of instantiators. Then, and only then, it will be reducible, and respectable.
  - g) I think first (with Fodor) that 'pain' is a bad example here. 'Belief that-p' would be better. But also Kim *seriously* underestimates how many possible disjuncts there are: possibly not only one per *individual* mammal-reptile-Martian, but one for each *at each time*.
  - h) Note that realizability talk *assumes* token-token identities. But Davidsonians about intentionality see intentional states as *vehicleless*, i.e. as only *globally* or (perhaps) *regionally* supervening on physical-physiological states.
  - i) The way I would put an objection to Kim cognate to the one Fodor makes is that he needs to add the *many levels* point to the *multiple realization* point. For pain or belief *does* take part in laws (just as pipe and filter do for functional concepts) *at that higher level*. They *are functional kinds*.
4. Three further arguments about supervenience and reduction:
- a) Beth's Definability theorem (proved from Robinson's Separation Theorem, and having analogs for various modal systems [van Benthem]):
  - b) Craig's reduction theorem:
  - c) Perverse reverse supervenience of physical (or anything vocabulary whatsoever!) on semantic-normative vocabulary. (This result does *not* depend on formulating things in terms of vocabularies rather than fact-kinds.)

### Part III : Kim "Multiple Realization..." and Fodor "Special Sciences" II

- 1. Kim (and Fodor):
  - a. Wonderfully clear summary of the rise of the consensus against reductionism based on the possibility of multiply-realized *functional* concepts-properties.
  - b. Argues that multiply realized functional properties, if thought of in terms of their realizations, would be seriously *disjunctive*.
  - c. Q: What is wrong with disjunctive concepts? Notice that you can't tell just *syntactically*: jade is a disjunctive concept, in the sense in which they are objectionable, and African emerald or non-African emerald is *not*. (But what about African elephant or non-African elephant? Here both

disjuncts are genuine kinds, and so is the disjunction. Fodor makes this point.)

- d. A: They don't play the right role in *laws*, hence are not *inductively projectible*.
- e. Kim claims that *if* psychological properties *are* functional and so multiply realized, then there cannot be a science of psychology.
- f. *Unless* it is a disjunctive science of the different sorts of instantiators. Then, and only then, it will be reducible, and respectable.
- g. I think first (with Fodor) that 'pain' is a bad example here. 'Belief that-p' would be better. But also Kim *seriously* underestimates how many possible disjuncts there are: possibly not only one per *individual* mammal-reptile-Martian, but one for each *at each time*.
- h. Note that realizability talk *assumes* token-token identities. But Davidsonians about intentionality see intentional states as *vehicleless*, i.e. as only *globally* or (perhaps) *regionally* supervening on physical-physiological states.
- i. The way I would put an objection to Kim cognate to the one Fodor makes is that he needs to add the *many levels* point to the *multiple realization* point. For pain or belief *does* take part in laws (just as pipe and filter do for functional concepts) *at that higher level*. They are **functional** kinds.

2.

- a. He opens with a very good summary of the prevailing orthodoxy (1992): multiple realization functionalist arguments convinced everybody that type-type reductionism could not work.
- b. He gives references to many of the classic articles relevant to this conviction.
- c. Note that I claim that functionalism in the philosophy of mind is one of the core programs of analytic philosophy, along with naturalism and empiricism.
- d. Say how functionalism keeps the virtues of both materialism and dualism about the mind: For the first, there is nothing non-physical involved, in the sense that all the realizations of functional states (e.g. valves) are purely physical. But it denies that all facts are physical facts, since there are also functional facts (e.g. about how valves go together with pipes, pumps, and filters).
- e. Discuss Quine's elephant topiary example, thought of now functionally, rather than behavioristically.
- f. There is the possibility that even within one subject, the brain-state that realizes a given mental-functional state may change over time. Thus what instantiates my thought that *p* may be different on different occasions.
- g. Note that Davidson and McDowell (among others: Lynne Rudder-Baker is especially good on this point) take beliefs, desires, and intentions to be states of the whole organism, that is, to be "vehicleless", to *have* no bearers or realizers that can be picked out as *parts* of the whole organism. This is a denial of token-token identity, of the Haugeland sort. It point to

the fact that “realization” talk carries with it token-token identity commitments.

- h. Note in passing that it is not clear that it is a good idea to put sentience and sapience in the same boat here. Vehiclelessness makes a lot more sense for sapient than for sentient states such as pain and other “raw feels”.

3.

- a. Kim summarizes Fodor’s argument as follows:

**argument for disallowing disjunctions. As I see it, Fodor’s argument in “Special Sciences” depends crucially on the following two assumptions:**

- (1) To reduce a special-science theory  $T_M$  to physical theory  $T_P$ , each “kind” in  $T_M$  (presumably, represented by a basic predicate of  $T_M$ ) must have a nomologically coextensive “kind” in  $T_P$ ;

**(2) A disjunction of heterogeneous kinds is not itself a kind.**

- b. He then develops the example of *jade* being disjunctive between *jadeite* and *nephrite*, pointing out that induction and confirmation won’t work with disjunctive antecedents, since all the examples may come from one disjunct, and there is then no reason to generalize to the other. If we allow arbitrary disjunctions, we get a *runabout inductive inference ticket* (to adapt Prior’s phrase about ‘tonk’).
- c. But on the multiple realization picture, ‘pain’ is the name of a second-order property: “the property of having some property with certain specifications” [MR 15]. So why isn’t it as bad as *jade*? If disjunctive properties cannot enter into the statements of laws, then it seems *pain* can’t.
- d. **This argument of Kim’s seems to me to overlook the possibility that what is disjunctive and non-nomic at *one* level is just fine as a nomic kind at *another* level. That is, it seems to me that we have here a situation where the *multiple realizability* argument needs to be *supplemented* with the *many levels* argument. For *functional* relations between *pain* and other concepts *at the same level* may be perfectly nomic. ‘Valve’ is realized in very heterogeneous ways, but the relations between valves and pumps and filters can remain constant and nomic across all of them.**
- e. Kim’s conclusion, by the end of Section VI, is “If MR is true, psychological kinds are not scientific kinds.” [MR 19].
- f. What he then wants to recommend is “multiple local reductions” of the mental to the physical, e.g. to human, reptilian, and Martian physiologies.

- g. BB: I think Kim underestimates how many kinds of local reductions he would need to do—certainly for belief. There might need to be not just one per individual, but one per individual *at a time*.
- h. Block objects that species-specific local reductions (and I am suggesting it is liable to be *much* worse) avoid the question of what all these have in common that makes them *pain* (or whatever). My argument in (d) addresses just this issue. Block says:

**ducible to physical states. Reduction, in the present debate, must be understood as reduction of types, since the primary object of reductive strategies is explanations and theories, and explanations and theories quantify over types....The suggestion**

- i. Kim concludes:

To summarize, then, the two metaphysical schemes I have sketched offer these choices: either we allow disjunctive kinds and construe pain and other mental properties as such kinds, or else we must acknowledge that our general mental terms and concepts do not pick out properties and kinds in the world (we may call this “mental property irrealism”). I should add that I am not interested in promoting either disjunctive kinds or mental irrealism, a troubling set of choices to most of us. Rather, my main interest has been to follow out the consequences of MR and try to come to terms with them within a reasonable metaphysical scheme.

I have already commented on the status of psychology as a science under MR. As I argued, MR seriously compromises the disciplinary unity and autonomy of psychology as a science. But that does not have to be taken as a negative message. In particular, the claim does not imply that a scientific study of psychological phenomena is not possible or useful; on the contrary, MR says that psychological processes have a foundation in the biological and physical processes and regularities, and it opens the possibility of enlightening explanations of psychological processes at a more basic level. It is only that at a deeper level, psychology becomes sundered by being multiply locally reduced. However, species-specific psychologies, e.g., human psychology, Martian psychology, etc., can all flourish as scientific theories. Psychology remains *scientific*, though perhaps not *a science*. If you insist on having a global psychology valid for all species and structures, you can help yourself with that, too; but you must think of it as a *conjunction* of species-restricted psychologies and be careful, above all, with your inductions.<sup>46</sup>

4. Note: Also point students towards the other two Kim pieces available from JSTOR, on Supervenience, and “The Concept of Supervenience”.



## Part IV a: Beth theorem

1. On the interpretation of the Beth theorem in this context:
  - a. In motivating the theorem:
    - i. Start with Carnap, giving a syntactic notion of *partial reduction formulae*. These are *not* biconditionals, but simply *conditionals* relating base-claims to target-claims (or predicates).
    - ii. These might be *very* thin, giving only a *few* connections. But we could ask, what if, without consisting of, or even *entailing* biconditionals that *define* the predicates of the one vocabulary in terms of the other, there were *enough* such inferential connections stipulated to *fix the meanings* of the target terms relative to the base terms?
    - iii. The only way to make this sense of ‘fixing the meanings’ precise is *semantic*: looking at what the terms are *true* of.
    - iv. That is what Beth does, in terms of *models* of the various predicates.
    - v. And his notion of implicit (semantic) definability turns out to be just the model-theoretic correlate of supervenience. It is the model-theoretic *correlate* only because of the difference between quantifying over *possible worlds* and quantifying over *models of a first-order theory*. And that is just the difference we would like to get clear on. (Cf. Etchemendy on the significance of this issue for our understanding of the model-theoretic conception of logical consequence. These are deep waters.)
2. On the Beth Definability Theorem:
  - a. Start with a *theory* T, which is just a set of sentences in a first-order language, and suppose that the *vocabulary* V of the theory (we can think of this as just the *predicates*, though the result will go through for individual constants too) is partitioned into two classes, a *base* vocabulary B, and the rest (V-B).
  - b. The terms of V-B are *explicitly definable* in terms of B iff there is for each n-adic predicate  $P_i$  in V-B a statement of the form  $\forall x_1 \dots x_n [P(x_1 \dots x_n) \leftrightarrow X(x_1 \dots x_n)]$ , where the only predicates appearing in X are from B.
  - c. But it may also happen that the terms in V-B are *implicitly definable* in terms of those in B. This happens [and this is the part I’m making up] when fixing the values of the predicates in B fixes the values of the predicates in V-B, i.e. the values of *all* the predicates in T. What is it to *fix the values* of a predicate? This is a model-theoretic notion: if we look only at models that satisfy the theory T, any models that agree in the interpretation of the B predicates (and individual constants) agree also in the interpretation of *all the rest* of the predicates (and individual constants) of T.
  - d. Then Beth shows that if T *implicitly* defines V-B in terms of B, then there are *explicit* definitions of the V-B terms in terms of B terms.

- e. And the point for us is that if V-B *supervenes on* B, then the conditions of Beth *implicit* definability are met, and so there are *explicit* definitions of the V-B terms in B terms.
- f. Kim appeals to this result, I think, to argue that one might as well insist on local reductions, rather than dropping back to (strong) global supervenience.
- g. In fact H&T's argument: The Beth theorem applies if and only if we restrict the models we consider to those that satisfy a given theory T. But the notion of supervenience that we want is that any two *genuinely possible*, or (here we have to be careful not to fall into a circle, and make supervenience trivially true) *scientifically possible* worlds cannot be alike in  $\phi$ -facts and not in  $\psi$ -facts. The class of worlds we want to quantify over cannot in general be taken to be those satisfying some first-order theory—at least not one that is even recursively enumerable. For instance, Gödel tells us that there is no such theory that gets all the truths of arithmetic in, and so all the worlds in which those arithmetic facts hold. They say [H&T 564]:

the major claims of physicalism. The principle of the physical determination of reference threatened to collapse to reducibility in view of Beth's definability theorem. However, as the work of Gödel and others would suggest, the power of our symbolic systems is such that full theoretical characterization of scientific possibility in any manner that would license the inference from determination to reduction is not to be expected.

The question is how this argument works.

3. There are the two related issues of what set of worlds we want to quantify over in making a supervenience claim, on the one hand, and the requirement of the theorem that they be all and only the worlds that model-theoretically satisfy some first-order theory.

- a. What I'll try to make of the Beth argument is the claim in (4) below. I think this is related to the Hellman worries about the appeal to first-order theories (though he thinks in terms of non-standard models, which Kim rightly suspects of being, in Kant's terms [Kim does not put it this way] *falsche Spitzfindigkeit*).
- b. But I think there may be a deep *connection* between
  - i. the issue of how to understand the relation between possible-worlds formulations of global supervenience and the model-theoretic notion of implicit definability that Beth appeals to, on the one hand, and
  - ii. my worries about the stability of the intermediate (Davidson-Sellars) recoil from mere supervenience, that looks to ontological naturalism (token-token identity) without ideological-reductive naturalism.

That is the line of thought sketched in (4) below

4. It is possible that the Beth theorem argument is the one I want to argue the instability of token-token ontological reduction of *objects* without ideological reducibility of *facts*, that is, the attempt to say that everything there is is physical, but that not all facts are physical facts.

- a. For the Beth result requires that our models have domains. Those domains are drawn from a super-domain, of everything there is. And predicates are interpreted by their extensions within that super-domain. It is looking at models of  $\psi$ -claims whose domain is the same as those used to interpret  $\phi$ -claims. For it is looking at *all* models of some theory T, no matter what the domain is. So, for instance, it is talking about models of  $\psi$ -facts that have, insofar as this makes sense,  $\phi$ -objects in their domains, e.g. a model of physical theory that replaces particles by thoughts.
- b. In connection with my discussion of this possibility—at the end of discussing the Beth argument [For now! We’ll obviously have to come back to this in Part III of the course.]—should put in my rant about the loose use of “mereological sums of particles (or particles-and-fields)” by physicalistically-minded philosophers like Lewis:
  - i. [Brief excursus on the made-up physics in which physicalist philosophical arguments are conducted:] For classical Newtonian mechanics, with gravity the only force, it was enough to specify the positions and momenta (which of course include the masses and motions) of all the (point) particles, to have fully specified the physical facts. Philosophers (like Lewis) sometimes still talk as though all we had to worry about was the ultimate constituent particles. With the addition of more forces (paradigmatically, electromagnetic ones), we needed to add more *dynamical variables*. These are *whatever* we need to associate with those particles to determine how their motions (and other dynamical variables) will evolve from one state to another. This gives us the idea of a *phase space*, defined by all the dynamical variables. Already with electromagnetic theory, it was not enough to associate values of dynamical variables with all the particles (Teilchen) or point-masses. We needed to associate them with all the points of space. Assigning values of dynamical variables to points in space is defining a *field*. So when physicalistically minded philosophers are being just a little more careful, they talk about “all the physical particles and fields.” This is OK, but substantially less definite in sense than is often supposed. For the magnitudes associated with points in some space (for QM, uncountably infinite-dimensional Hilbert space) need not be simple *scalar* magnitudes (which would include probabilities), but can include vectors and higher-order tensors, or indeed, *any* algebraic structures. Since the space the structures are mapped onto need not be ordinary space-time, at this point it is not clear what is being ruled out. Talk of ‘fields’ is in danger of becoming just talk of mathematically characterizable

structures in general, since almost any such can be horsed around into the form of such an abstract field.

- c. The way the model-theory uses the domains amounts to a kind of token-token identity assumption built into the apparatus of model theory itself. Notice that Haugeland's loop and arrow example depends for its cogency on the assumption of a base domain common to both representations. The challenge is to conceive of the case where there *is* no such common base.
- d. Nonetheless, it may be that every physicalist(naturalist)-ontology-without-physicalist-ideology satisfies this part of the presuppositions of the Beth theorem. Note that this consideration would *not* show that they must satisfy the *other* basic assumption of that theorem: that we are looking at all the worlds that satisfy (in the model-theoretic sense) some *first-order theory*.