11/28/2006

Naturalism—Week 11 Seminar Plan:

1. Plan:
    a. A recap of the argument and conclusion from Week 10, about the instability of the proposed *via media* between reductionism and global supervenience that consists in "token-token identity theories". Modal separability argument, sortals, contingent identities, and the genus of which material constitution, functional realization, and the relation between human beings and persons are species. [I need both a name for and a characterization of that genus of relation between things with different criteria of identity and individuation.]
    b. Then the bits from the end of the Week 11 notes that I did *not* get to. ((22) and (23); add (21) to discussion of trans-sortal identities, H_2O…)
    c. McDowell, including my questions about the threat of a new dualism (if he can't explain the *relations* between the distinguished items, i.e. in a slogan, if he needs a descriptive/normative pineal gland).
    d. To underline the questions McD does *not* address: the Paradox of Mechanical Rationality, and GOFAI response (early Fodor).
    e. Issues that raises about vehicles; 2 senses of vehiclelessness and holist and normative arguments for them; trans-sortal identities in that context; holistic role abstraction.
    f. Then a gesture at pragmatic naturalism: Rorty, Price, Ramberg, and bits of my BSD:TAP, with its synthetic expanded-nature pragmatic naturalism— gotten by being an expanded-nature naturalist about the vocabulary in which one conducts one's pragmatic naturalism. On this way of drawing a moral, Price is both too radical, in rejecting representation-description as part of his pragmatic naturalism, and not radical enough, in being a first-nature naturalist about the vocabulary in which his pragmatic naturalism is conducted.

2. The modal separability argument:
    a. The modal separability argument does not show that there cannot be contingent identities. That is, it can be the case, for instance, that $(\iota x Dx = \iota y Fy)$ & $\Diamond{\sim}(\iota x Dx = \iota y Fy)$. They will all be:
        i) Cases where the sortals have the same criteria of identity and individuation, but possibly different criteria of application, as in "the big lion = the mammal in the cage," and "the tall woman = the most important person in the room." (Note that this much is true of "this cat = that mammal," and "this woman = that person," but that these are *necessary* rather than contingent, because the demonstrative rigidifies the specification, and so makes them *de re*. See (iii) below), or
        ii) Cases where one of the sortals is *suitably schematic* (see below).

iii) *De dicto*: ιxDx, *whoever or whatever that turns out to be* = ιyFy, but in some worlds the thing that satisfies the first description may not satisfy the second. If we first pick out the *people*, say, and follow *them* through other worlds, *they* will be identical in *all* worlds (in which they exist at all).

b. It does show that there cannot be contingent identities between items that fall under sortals that have different criteria of identity and individuation (except in the special case where one of the sortals is suitably *schematic*).

c. Contingent identities always involve things that fall under the *same* sortal. What is contingent is that their criteria of *application* coincide, though they need not. That their criteria of identity and individuation coincide is a *necessary* truth about the thing(s).

d. Thus "the first Postmaster General" and "the inventor of bifocals" may or may not pick out the same individual. But whether they do or not, it is a necessary truth that both pick out *persons*.

3. The modal separability argument tells against *most* (see the conditions above) trans-sortal "token-token" (recall the reservations I expressed last time about using this label generally) contingent identity claims. The objection is *not* that such claims are *unintelligible* (as Putnam suggests). That is true of some that involve seriously *schematic* sortals, but it is not generally true. It is that the MS argument shows, via the indiscernibility of identicals, that such claims are *false*.

a. There is a discussion at the end of these notes about how to understand some classic cases of such reductions: "water = H₂O", "lightning = atmospheric electrical discharge." For now, just notice that they are one and all put forward as *necessary*, *type-type* identities, not contingent, token-token identities.

b. But one important case that the argument does *not* seem to apply to is Davidson's anomalous monism. That is because Davidson's theory of events—whether consciously and intentionally constructed so as to avoid this sort of difficulty or not (he doesn't ever say anything about the sort of considerations one which the MS argument relies)—has very special features that render it immune to the modal separability argument.

c. In particular, he denies that one can re-identify an *event* if *anything* about it changes. *Every* description under which they fall is *essential* to events. That is, events as DD construes them *have no counterfactual properties*. There is no way to pick out *that very event* in *any* other possible world. In a different world, there is only some other event, which may correspond to the original one in various ways, but cannot be *identified with* it.

d. Query: It seems that he *must* exclude *intentional* descriptions, in the sense not of the specifications under which the events that are also actions *are* intentional, but in the sense of those that are formed from propositional attitude ascriptions. Would the Second Punic War really have been a different event if I had never heard of it? But DD seems to offer no such exclusion.

e. This feature is a feature of Davidson's account of events that is unusual, even bizarre, and it has not come in for a lot of discussion. Does it really make sense to think of there being a kind of thing, events, that has this very peculiar ontological property of being welded into its world in this way? I suppose so, for all we need to create such a category is to take any sortal K, with its criteria of identity and individuation, and its criteria of application, and form from it another sortal, K*, and stipulate that the CofI&I of K*s include those of K (as 'passenger' includes those of 'person') and in addition the condition that *a* and *b* must share *all* their *actual properties*, including their *relational* in order to be identical K*s. Such objects will be *parochial* to their worlds, able to exist *only* in *that* world. Davidson just defines events so that they are *modally parochial* in this sense.

f. That does avoid the argument, via modal separability and the indiscernibility of identicals, against contingent trans-sortal identities of mental events with physical events, of the sort that anomalous monism asserts.

g. But at what price? What now is the difference between this sort of "token-token contingent identity" claim and a claim of mere *global supervenience*? What surplus meaning is there to the token-token identity claim over and above the claim that any difference in the *mental* events between two worlds entails a difference between the *physical* events in those two worlds? (Compare (26) below.)

h. This is not a rhetorical question. There may in fact *be* a residual difference, which makes the anomalous monism claim stronger than just supervenience. I haven't thought enough about this issue to be confident either way. The first place to look is probably the arguments Horgan retails (really taken over from Kim) for thinking global supervenience is too weak to be interesting. One of them, the total lack of *explanatory* connection, is shared by anomalous monism—that is the 'anomalous' bit. The other, about tiny, distant physical differences being enough to make *huge* difference (between mindedness and not), compatibly with global supervenience, is what *ought* to be ruled out by a token-token identity theory. But is it?

4. (1a): After the instability of the proposed *via media* between reductionism and global supervenience that consists in "token-token identity theories":

a) So where do these arguments leave us? We can say that persons and intentional states *globally supervene* on (first) natural objects and states—though we haven't seen knock-down arguments even for that. But we have also seen that to say that is not to say very much.

b) We *can* say something more, however. At least *one* way that particular natural objects can be associated with singular terms whose sortals exhibit different CofI&I is as *vehicles* for *holistic role abstraction*, taking us from ground-level objects-and-relations to upper level ones. (Cf. (4) and (3) above.)

c) The big question then is: what *relations* among *ground-level* objects, both objects and relations being specifiable in *naturalistic* vocabulary (however we demarcate that category, whether narrowly or in expanded terms), is it that induce the holistic role abstraction?

d) My answer in *MIE* is that it is *social* relations of being *taken* or *treated* in various ways, and furthermore, that it is those relations specified in a *normative*, but not necessarily *intentional* vocabulary.

e) We could, at any rate, envisage a *hierarchy*:
    i) physicalistic vocabulary;
    ii) core natural-scientific vocabulary;
    iii) that plus special sciences (in a way, this is where both Millikan and Fodor are w/res to naturalistic base vocabularies);
    iv) *social* sciences, but still descriptive-explanatory, in a narrow sense;
    v) social sciences, but read more broadly: as Geisteswissenschaften, in a sense that includes *normative* vocabulary.
    vi) Social sciences, read still more broadly, as Geisteswissenschaften in a sense that includes also *intentional* vocabulary.
    vii) Geisteswissenschaften in a yet broader sense, which includes also *hermeneutic-interpretive* vocabulary and methodology. [Note that this may or may not be in principle distinguishable from (vi).]

f) Whatever choice is made here, one need *not* (and if the modal separability argument goes through, *can*not) see the relation between the *holistic roles* that result from the abstraction (in a literal, technical sense of constructing equivalence classes) as one of *identity*, hence not as a *reductive* relationship. It is a special kind of vehicle relation: that whose paradigm is the relation between *pen-and-ink scratches* and *signatures*.

g) It is a *much* more specific relation than mere global supervenience.

h) It is a kind of *functional realization*. But at least once we are at level (e-iv) and above (as Millikan, but *not* Fodor, is), it is a *social* functionalism, and at (e-v) and above, a *normative* social functionalism.

i) These relations between distinctly sortalized individuals support *some* definite kinds of *explanatory* relations across the levels. For some counterfactuals regarding ground-level facts have consequences for upper-level facts, and *vice versa*. [Examples:]

5. (1b): <u>Model theory in terms of domains vs. modal theory in terms of possible worlds</u>:

    a. The domains of the relational structures that we deal with in model theory consist, we are told, of *objects*. They are distinct from one another, and re-identifiable, but nothing is said about what sortals provide the criteria of identity and individuation. They are distinguished independently of and in advance of the definition of any *properties* of or *relations* between them. (The properties and relations are represented by sets and sets of tuples of those domain objects.) Domain elements can be *counted*, and there are facts about whether we have picked out two different ones or the same one twice. Yet, as we have seen, 'object' is a purely schematic sortal, which provides no CofI&I capable of underwriting all these claims. And

'element', even 'domain element' is no better, except insofar as it picks things out as members of a certain set. But the members of a set must be identified and individuated prior to defining the set, and the question we want to think about is how that is supposed to have been done in the case of the members of the set that is the domain of a model-theoretic relational structure. (Recall the earlier observation that even *numbers* are not bare particulars that are "merely numerically distinct." Numbers of all sorts have CofI&I defined in terms of their order properties and relations to one another.)

b.   The only way this can work, as far as I can see, is that it is implicitly assumed that in the metalanguage in which we specify the models, we have at our disposal genuine, non-schematic, sortals sufficient to identify and individuate the members of the sets that are our domains, and hence the elements of those domains. The point then is that for the purposes of the work we will do with the resulting relational structures in our model theory, it doesn't matter at all *what* those sortals are. Since it does not, we can just talk about the elements as *objects*, i.e. as *merely* identified and individuated.

c. So there is a sense in which model theory can do without sortals, and a sense in which it cannot. The sense in which it cannot—that the metalanguage in which we specify the models must deploy genuinely individuating sortals—is philosophically important in various contexts. It means, for instance, that we cannot think of the semantics of *all* of our languages in purely model-theoretic terms. For model-theoretic semantics is *parasitic on* languages with individuating sortals, which can accordingly serve as metalanguages for specifying domains and relational structures. Of course, once the enterprise is off the ground, there is nothing to stop us from giving a model-theoretic account of sortals and sortal predicates. But that possibility arises only downstream from the specification of models in the first place.

d. The sortals that get the model-theoretic enterprise off the ground to begin with remain implicit, offstage, functioning as ladders we can throw away after we have climbed up with them. By contrast, in the case of possible worlds we should keep in mind Kripke's point that possible worlds should not be thought of as things we look on as through a funny kind of telescope, unsure how to identify what we find there. Rather, we *say* what possible world we are talking about. Saying what objects are in it requires using genuine, non-schematic, individuating sortals, or expressions (such as proper names) that have such sortals associated with them. I *stipulate* that the world I am talking about is one which *Ben Franklin* did not invent bifocals. And in saying that it is *Ben Franklin* I am talking about, I commit myself (whether or not I fully understand the content of that commitment) to the criteria of identity and individuation that go with *that* person. In this setting the sortals are front-and-center, explicit and on-stage, not implicitly presupposed as part of the antecedent activity of setting the stage.

6. (1b): <u>World of facts vs. World of objects</u>:
   a. If there really is "no entity without identity," then a world of *objects* must come with *sortals* specifying their CofI&I. So they cannot be *bare* objects, as in model theory—*merely* distinct. (Note that even *numbers* are not "merely numerically distinct." They *always* come with *order* properties and relations.)
   b. And the CofI&I are specified in terms of various *adjectival predicates/properties*. So *because* they are sortalized, the objects will also come with at least *some* properties (descriptions true of them).
   c. This is all moving them in the direction of *facts*. That is, the notion of a world *merely* of objects is incoherent.
   d. *Stipulating* a world of *basic objects* (say—madeup philosophers' physics alert—particles-and-fields) and mereological sums of them, as *all* the objects in a world—i.e. stipulating that *every* object in the world is *identical* to one of *these* objects—is, whether one knows it or not, stipulating that there are *not* certain kinds of objects in the world: biological objects, or intentional-hermeneutic ones. (At least, this is true *if* the argument from modal separability for distinguishing <u>identity</u> from <u>constitution</u> holds up.)
   e. The same question will come for any world-of-objects fan (a view that Kotarbinski called 'reism': that all there is, ontologically, is things of the category of *thing*, res) that comes up for model theorists: How are the ˢobjectsˢ identified and individuated? *Some* sortal must be being applied, at least implicitly, *by us* when *we* talk *about* this world of objects. For there is no entity without identity, and identity is intelligible only in the context of determinate individuating sortals. And 'object' is not such a sortal, being purely schematic. It seems that we need at least some properties and relations of those objects, for the criteria of identity and individuation to appeal to. And that means that there *will* be facts *about* those objects, as well as the objects themselves.
   f. Indeed, it seems that the necessity to individuate the objects by specifying associated individuating sortals provides an *argument* for the *priority* of the world as everything that is the case over the world as a collection of objects.

7. (1c): McDowell's "Naturalism in the Philosophy of Mind": Index Sellars passage and McD on the space of reasons [mention the nickname of our seminar room]:
a) The Sellars passage is written as though the distinction between the space of description (and explanation, hence where modal vocabulary applies) and the space of reasons (articulated in normative terms) were one of mutual exclusion.
b) The fact that characterizations of something as a knowing or a believing are declarative sentences that embed properly seems to militate against an account exclusively in terms of what one is *doing* in offering one: namely, *placing* something in the space of reasons and justifications. For that sort of pragmatic force is stripped off by embedding.

c) The conclusion (cf. post Frege-Geach ascriptivist-expressivists like Gibbard and Blackburn) seems to be to restate Sellars's claim: "In characterizing an episode as one of *believing* [since the difference between this and *knowing* is *not* to the point he is making], one is not **only** describing it. One is **also** placing it in the space of reasons, of justifying and being able to justify what one says."

d) So 'belief' *does* have a descriptive content, and in applying it, one is, *among other things one is doing*, describing it. But to understand the *content* of that description, one must understand what one is *doing*, *besides* describing it, in placing it in the space of reasons.

e) What (else) one is doing (besides describing it) is taking up a normative stance: attributing a commitment to the believer. But one is *also describing* the individual *as* committed, *as* having a *normative* status, namely being committed.

8. Draw together threads from previous weeks:
   a. Sellars on calling something a knowing not being *describing* it, but placing it in the space of reasons.
   b. This means that the same is true of *believing*.
   c. So Sellars is contrasting *describing* with *putting in the space of reasons* (the last being the term McD uses to distinguish what he cares about from the realm of laws).
   d. What is the relation between *describing* and *representing*?
   e. Isn't one *describing* and object, *representing* it as being some way, if one applies a *sortal*? That is, having the sort of inferential involvement exhibited by criteria of application (and consequences of application) and CofI&I is *sufficient* (and may be necessary) to distinguish *something* (maybe description) from mere labeling. But if we are not to be *declarativists* about description (which will rule out Sellars' point, and maybe McD's, since normative vocabulary will be descriptive), it seems that falling under a sortal should be *sufficient* for being *descriptive* of objects (which does not rule out that only some descriptions *also* have a *normative* component, and that *some* normative expressions may have *no* descriptive content).
   f. 'Belief' is a sortal.
   g. The Sellars point is that 'belief' is a *normative*, not (at least exclusively) a *descriptive* sortal.
   h. That could be because the criteria of application are normative, or because the criteria of identity and individuation are normative.
   i. A sufficient condition of *a* and *b* being different beliefs, because they would then be different *kinds* of belief, is that they have different *contents*, as specified in the 'that' clauses. (Which is not to say that every difference in the 'that' clauses expresses a difference in content.)
   j. The contents are individuated in part by their relations to *other* contents.
   k. Evans's "Generality Constraint" says that for one to be able to have the belief that Pa, one must be able to entertain Qa, for at least some Qs, and Pb, for at least some bs. Fodor accepts this. That is why he thinks that there must be vehicles that vary in their non-intentional properties in a way that is isomorphic to the content differences expressed.

l. The *modal separability* argument says that any object can *be* (identical to) a belief only if it is individuated the same way.

m. Besides the generality constraint, Quinean considerations dictate that to be able to *entertain* various contents, one must have other particular *beliefs*.

n. (j), (k), and (m) together entail that the individuation of beliefs, according the CofI&I of that *sortal*, must be *holistic*, in the sense they define.

o. So *if* we try to identify a belief with some *sub-personal* state, some state that one can be in *independently* of other states one is in, the *modal separability argument* will show that that state cannot be *identical* with the believing. This shows that beliefs must be *vehicleless* in *one* of Kris's two senses. For the *only* way to ensure (at least, a *necessary* condition) that the modal separability argument does not do in the identity, given holism about the individuation of content (and that content differences are sufficient to distinguish beliefs), is that we look to *total* states of the believer. These must *at least* be states of the whole person, and *perhaps* even states of the whole *world*.

p. But we can *also* see an argument from (the demands made by considering) *modal separability* to the *second* sense of 'vehiclelessness'. For the individuation of contents is not just *holistic*, but *normative*. It is a matter of what *properly* follows from, is evidence for, and is incompatible with what. If that is right, then nothing individuated non-normatively can end up being trans-sortally identical with beliefs (believings).

q. Here my argument about the *normative* character of *intentional* explanations comes to the fore.

r. We are now combining *modal* considerations, having to do with *explanation*, with *normative* ones, having to do with the *space of reasons* (cf McD).

s. For Sellars, description and explanation go hand in hand.

t. And modality makes explicit *explanatory* relations.

u. But there is *explanation* in the space of reasons, too.

v. But it is *normative* explanation, specifically *intentional* explanation.

w. Normative vocabulary : Intentional Explanation :: Modal vocabulary : Explanations of descriptions.

9. (1c) McDowell: McD [Continue through the handout of passages, discussing each in the right place, as part of general exposition of the article.]:

a. First-nature naturalism vs. second-nature naturalism;

b. McD must reject, or overcome, a number of distinctions (moving to a higher genus): nature and culture, the natural and the hermeneutic forms of understanding (Erklaerung vs. Verstehen, Natur- and Geisteswissenschaften), things that have *natures* vs. things that have *histories*. Note that it was an essential element of high-church unity-of-science physicalism to insist that there is only one sort of understanding, and that natural science is the measure and the result of it. So there is a pleasing symmetry to McD's position—not to be confused with the symmetric reciprocal containment assertions of Heideggerian-Romantic inclusion of scientific image in the manifest image, as one more "form of

life". (We'll see just below that Wittgensteinian-classical-pragmatist anthropological approaches—Darwinian, as it were, from *above* rather than *below*, Hegelian in the Deweyan key—belong under a different heading.) What McDowell does is absorb the first-nature natural "realm of laws" into a notion of nature that has been expanded to include second nature.

c. Here we should ask whether there is something corresponding to *description* that goes with this new-and-expanded notion of *explanation* (cf. Sellars on description-and-explanation) (namely Verstehen, hermeneutic explanation, which would include *intentional* explanation, with its connection to practical reasoning, as well as broadly *textual* interpretation—which it would be inappropriately reductive to assimilate to practical reasoning more generally (as opposed to seeing the hermeneutic as including that sub-species), particularly, but not just in this case, when practical reasoning is constricted by being identified with its *instrumental* species). And it seems that there does exist such a corresponding kind of description. It is *intentional* description, description in terms of *reasons*, doxastic and practical *commitments*. In fact, this sort of *description* is to the intentional-hermeneutic as the *modal* is to the empirical-descriptive as the *normative* is to the *modal*. Where first-nature naturalism incorporates the *modal*, second-nature naturalism incorporates the *normative*. (But think how Jackson would think about just redefining the descriptive to include the normative, in order then to vindicate a kind of descriptivism about the normative. Surely he would mutter something about such a strategy having all the advantages of theft over honest toil.)

d. On the complaint in (b) that McD is committed to rising to a higher generic synthesis of the two sides of various important distinctions (so— he could say—ensuring that these *distinctions* do not become *dualisms*) he might well invoke the fine, if old-fashioned, notion of <u>natural history</u>. The thought would be that talking, culture ("Sprache is the Dasein of Geist,") is an episode in the natural history of our kind of animal: human beings.

Quotes from and Questions for McDowell:

10.

a. "This is the result of a familiar trade-off; the price of discarding Cartesian immaterialism, while staying within restrictive naturalism, is that one's singled-out part of nature is no longer special enough to be credited with powers of thought. But Millikan's conception, for all its freedom from immaterialism, is like the original Cartesian conception in threatening the sane belief that a *res cogitans* is also a *res dormiens,* a *res ambulans*, and so forth. Millikan's "intact mind" does not exercise rationality, and the "head-world system," which does exercise rationality, is not the thing that sleeps and walks. The rational animal finds no place in the picture." [104]

b. McDowell's *liberal* naturalism, as opposed to *restrictive* naturalism (what I've called "expanded-nature naturalism") precisely includes sortals such as 'animal', with *biological* criteria of identity and individuation that preclude the *identification* of *any* animals with (mereological sums or structured constellations of) their material constituents. There is no possible world containing only a single wombat. The Anscombe-Thompson-Rödl notion of <u>logical form</u> presents a further articulation of this thought, one aspect of which is reflected in the modal separability argument against a large class of trans-sortal "token-token" identity claims.

c. McD makes a remark about the "head-world" system that, while literally true, seems to me a cheap shot as expressed: "But 'externalism' is a grotesque if it implies that exercising semantic rationality is an activity of a "head-world" system—as if the environment of what we ordinarily conceive as thinkers is partly responsible for doing the thinking that gets done." [103] That *is* grotesque, but it is *not* the thought Millikan (like others) is endorsing. The idea is rather that what the "intact mind", thought of in *non*-intentional terms genuinely does *counts as* thinking *only* in the larger context. That thought is no more grotesque than the idea that making marks on paper counts as a signature, or as entering into a contract, only in the context of a larger system—as though it were not *I* who take out the mortgage by that act, but only the whole system. The good thought McDowell is after here is the one made more precise by the modal separability argument: one is *not*, by these means, going to be individuating *thinkers* in the same way one individuates *animals*.

d. I pointed out last time that the modal separability argument, in the context of certain other auxiliary hypotheses, would put pressure on the idea that we (persons, believers and knowers, rational agents) *are* animals, or human. Given those same collateral premises, it would threaten "the sane belief that a *res cogitans* is also a *res dormiens,* a *res ambulans*, and so forth." But what else one needs is the claim *either* that *this very person* could exist *without this very animal* existing, or *vice versa*. And McDowell means to be denying that. *We*, persons, rational agents, the subjects of normatively individuated intentional state- and episode-attributions, *are* (identical to) human animals. The common individuating sortal is a *biological* sortal. We need not think that it is 'mammal'—for that would commit us to the impossibility of non-mammalian persons. But McD is committed to there having to be *some* sort of specifically *biological* creature that each person is identical to. If inorganic, electronic computers could be persons, it would follow, for McD just as according to the modal separability argument, that *we* are *not* animals (and so, *a fortiori*, not *rational* animals, or sleeping or walking ones either).

11.

a. "Of course there is a relevant organ, the brain, and none of what I have said casts doubt on investigating how it works. But on pain of losing our grip on ourselves as thinking things, we must distinguish inquiring into the mechanics of, say, having one's mind on an object from inquiring into

what having one's mind on an object is." [104—penultimate 'graph of the essay]

b. This passage seems to make the distinction I put in terms of that between saying what doing the trick consists in and saying how the trick is done— how to go about doing it.

c. But has McD left room for an investigation into the *mechanism* of having one's mind on an object?

d. He has earlier said: "Millikan takes the introduction of sense to be a tool for characterizing the mechanics of having one's mind on objects. She rightly concludes that grasping a sense, so construed, cannot be what Frege wants it to be, but she does not see that this might tell against the construal." [102] McD's view seems to be that the task for which the notion of sense is introduced is not this, but rather saying what it is to have an object in mind.

e. What seems most to be missing from McD's comments here is any view about how to understand the *relation between* an account of "the *mechanics* of having one's mind on an object" (having an object in mind, being able to think *about it*) and *what it is* to do those things. We can agree that these are *different* undertakings. And we can agree that, therefore, they should not be *confused* with one another. So we can agree that "If the only respectable intellectual orientation toward rationality is inquiry into how it is 'effected' in a mechanism, we lose our grip on rationality as something exercised in the activities of an animal." [104] But if *both* are respectable enterprises, the question still remains how they should be understood to be related.

f. McD generally wants to reject "How possible?" questions in philosophy, preferring to replace them by diagnostic exercises intended to dispel the confusions that underlie our temptation to find something mysterious or in need of explanation. But it seems here that the paradox of mechanical rationality gives some point to the "How possible" question.

g. Among the things we want to know is how any mechanism *could be* a mechanism for "having one's mind on an object," once we have granted that that characterization is not a description of something in non-normative terms, but involves placing it in a normative space of reasons, rational reponsibilities (with justificatory, ampliative, and integrative dimensions). How is it possible for something to be a *mechanism* for acquiring such a *normative status*?

h. The notion of holistic role abstraction provides the form of one possible response that is different from both the reductive and the epiphenomenalist moves. But it leaves the question as to whether the *base* relations, from which one is to abstract roles, can be characterized in *non*-normative terms, if the results are to be recognizable as *conceptual contents*. [This remark foreshadows the discussion later, of HRA as a response to the paradox of mechanical rationality.]

i. McD's view is that "The proper home of the idea of 'grasping senses' is in *describing* patterns in our lives—our mental lives in this case—that are

intelligible in terms of the relations that structure the space of reasons….Liberal naturalism needs no more, to make the idea of 'grasping senses' unproblematic, than a perfectly reasonable insistence that such patterns shape our lives." [103] Emphasis added. Note that he is happy to talk about *describing* those patterns, even though they are patterns in the space of reasons.

j.   But is it enough to make the notion unproblematic that we show it has a place in space-of-reasons talk? The *distinction* between causal-explanatory-descriptive talk and intentional-explanatory, space-of-reasons talk is OK in the sense that we can be pretty clear about how each sort of discourse works on its own. But, in accord with my slogan that a *distinction* becomes a *dualism* when it is drawn in terms that make the *relations* between the distinguished items unintelligible, we must ask whether McD is not just asking us to acquiesce in a dualism here. For he seems unwilling to say *anything* about the *relations between* first and second nature, between the two sorts of intelligibility or explanation.

12.

a.   One issue, among the many that he does not address concerning the relations between the "realm of laws" and the "realm of reasons", is whether the relations between them should be made intelligible in first-nature naturalistic terms of causal explanation and laws, or in intentional-hermeneutic terms of reasons and vocabularies.

b.   It seems to be the upshot of his discussion that we can*not* understand the realm of reasons in the terms of natural science: "The fundamental mistake is not the notion of a ghostly mechanism, but the idea that the mental can be in view from a standpoint that organizes its subject matter in the manner of the natural sciences." [99]

c.   That seems to leave just two options. One is that we understand the relation between these realms of objects in terms of the relations between the two *vocabularies*—in the idiom of my contribution to the Rorty volume, by using the vocabulary vocabulary rather than the causal vocabulary. This strategy could be worked out either in the Heideggerian-Romantic form, or in the pragmatic naturalist form.

d.   The other option would be that a *third* sort of intelligibility is required to address the *relations* between the first two modes of intelligibility. (Perhaps a distinctively *transcendental* mode of intelligibilitiy?)

e.   What I don't see is that McD can be entitled simply to put this issue to one side. Doing that is what I called above "acquiescing in a dualism"—now not the cartesian dualism, but the kantian one, of causes and reasons, or facts and norms.

13. (1d), Paradox of Mechanical Rationality: The Paradox of Mechanical Rationality is one challenge it seems McD is obliged to respond to:

a.   Paradox of mechanical rationality—or physical rationality, or first-nature-naturalistic rationality, as per Haugeland's *Artificial Intelligence: The Very Idea*. Intentional explanation (and hermeneutic understanding, if that

is different) depends on the efficacy of *meanings*, *conceptual* or *propositional contents, reasons*.

b. On the one hand, only if something is intelligible as *sensitive to* such items, features, or properties can it be described as *rational* (or, if, as per Sellars, this is not *describing* it, only in this case can rationality be *ascribed*).

c. On the other hand, first-nature naturalistic vocabularies do not mention these items. Physics does not acknowledge the efficacy of meanings. A physical system is one that is *in some sense* sensitive *only* to purely physical forces.

d.   So: how can a physical system be rational?

e. In McD's terms: how can we understand the possibility of something with a *first* nature *acquiring* a *second* nature? We may admit *that* it is possible. The challenge is to understand *how* it is possible.

14. (1d): One possible answer is *reductive*: meanings (reasons) just *are* a particular constellation of causes. Fodor's claim that representational content is a matter of asymmetric counterfactual dependence of representings on representeds ('horses' on horses) would be a paradigm. On this account, meanings are found within the causal realm, as items that *can* be causally efficacious in the way other physical causes are. For if representing horses is just a matter of being appropriately "nomologically locked" to horses, then the *fact that* something represents horses (is a 'horse') *can* have causal effects on other physical items.

15. (1d): Another possible answer to the "How?" question is that of classical AI (what Haugeland calls "Good, Old-Fashioned AI", or GOFAI):

a. Semantic contents always have *bearers* or *vehicles*.

b. Those bearers or vehicles are just physical items: marks on paper, noises, or neurophysiologically characterizable states.

c. *If* they vary systematically (cf. Evans's Generality Constraint) and be related to one another (cf. Quine's Meaning-Belief Constraint) in ways that are *isomorphic to* the way meanings vary systematically, and are related to one another and to beliefs having those contents or meanings,

d. *Then* physical interactions of the vehicles ("tokenings") will be just *as if* there were sensitivity to reasons, contents, or meanings.

e. Haugeland's characterization of the strategy: "Take care of the syntax, and the semantics will take care of itself."

f. This is how calculators "manipulate *numbers*", which do *not* exert physical forces. They do that by manipulating *numerals*, which *can* exert and be influenced by physical forces.

16. (1d):

a. Notice that this formalist response—as we might call it, since it insists that *isomorphism* is all that is required for 'sensitivity' to meanings—is an *epiphenomenalism* about the semantic. All there *really is* is stuff *isomorphic* to the semantic.

b. And we might ask: Could it really be like this *all the way down*? That is, does it make sense to say that this sort of causal manipulation of vehicles

in a way that preserves an *isomorphism to* the semantic is the *only* sense in which *anything*, including our own thoughts, *has* a content?

    c. One possible response to this worry might come from Davidson's measurement metaphor (or model): we use 'that-p' clauses to *measure* the attitudes of others, mapping their noises onto our own uses of sentences, in order to offer intentional explanations of them (render them intentionally intelligible). This, without that metaphor, is also how Dennett thinks of things.

    d. This *formalist* response to the "How?" question is the one Searle rejects in "Minds, Brains, and Programs," where he pumps the intuition that I could do all of the in-fact-isomorphic-to-semantics manipulations, and still not *understand* anything, not *grasp any meanings*, not *feel the force of any reasons*.

17. (1e) Formalism and vehicles:

    a. The formalist response requires vehicles.

    b. And those vehicles must have non-intentional specifications.

    c. Under some of those non-intentional specifications, there must be an isomorphism with the relevant semantics sufficient to ensure that any system of things standing in those relations to each other can count as *expressing* the meanings in question.

    d. This means that there must be systematic variations in the non-intentional specifications (descriptions) of the vehicles corresponding to the semantic variations (paradigmatically of terms and predicates), in order to satisfy the Evans Generality Cosntraint.

    e. And it means that there must be a difference between representings that are *endorsed* and those that are not (there must be a "belief box"), in order to satisfy the Quinean Meaning-Belief Constraint.

    f. This last is what Sellars is putting pressure on in his dictum that in calling something a *belief* one is not describing it. For one is rather taking up a certain sort of *normative* stance towards it: attributing a *commitment*, a *responsibility*—something that *answers to how things are* in the sense of being *correct* only if things satisfy a certain description. (Of course, in describing something, one is *also* taking up a normative stance. In *calling* something a description, one is doing something *different*, taking up a *different* stance: not making oneself answerable, but taking someone else to be answerable. In *MIE*, these two stances, one first-personal and the other third-personal, are seen as two sides of one coin, two sorts of social-perspectival deontic attitudes that in principle make sense only as part of a constellation that includes both.)

    g. It is a substantive claim, however, that one can characterize sufficient conditions for expressing conceptual contents in these formalistic terms. What works for numbers and numerals may not work for *empirical* contents. For those it seems we need at least language-entry moves, and possibly also language-exit moves. And there we have again the question of whether they can be characterized in purely causal-descriptive terms, or whether *normative* vocabulary will be required.

18. Both of the arguments suggested above, connecting the modal separability argument concerning identities between intentional sortals and non-intentional ones (the arguments from the holism of intentional content, and the argument from the normativity of the connections between intentional contents) are delicate, and require being set out with considerable care—as I do not above.

19. Argument from holism to no sub-personal state being identical with belief:
  a. On the Davidsonian picture, there is a whole-person state that is believing B, where B is a whole total-belief-and-desire constellation. The belief that p can only be identified with a disjunction of those total states: where it is true of each disjunct that it includes the belief that p, and there is a disjunct for *every* total intentional state that includes the belief that p.
  b. The idea of the argument in question here is that if one identifies the belief that p with anything less than a total-person state, then it will be *independent* of *other* such less than total-person states, in a way that violates the Evans Generality Constraint or the Quinean constraint that the capacity to mean that p depends on having various collateral beliefs.
  c. As to the first, Fodor's picture of isomorphically varying vehicles seems sufficient to guarantee that the Generality Constraint can be met.

20. Argument from normativity of individuation of intentional contents to intentional states having no non-intentional specifications [cf. (24) and (25) below]:
  a. This one is pretty straightforward, I think (at least, relative to the holism-personal-level one).
  b. We'll look at the two sortals involved: the intentional-state sortal, 'belief', and the non-intentional specification sortal. More specifically, we'll look at the criteria of identity and individuation associated with those two sortals.
  c. The modal separability argument says that if there are *any* counterfactual circumstances in which those two sets of CofI&I would diverge, then *no* item falling under the one sortal can be *identical* to *any* item falling under the other.
  d. So the issue comes down to whether one can reproduce exactly the normative connections between intentional contents in non-normative, non-intentional terms.
  e. One can*not* do that (arguably) if what we have on the non-normative (hence non-intentional—on the assumption that intentional contents are essentially normatively individuated, that is, individuated in part on the basis of what inferential moves would be *correct*) side is mere matter-of-factual *regularities*. What we would need to argue is that we could always come up with a case where acting so as to fulfill the regularity would not be correct, or what would be correct would not be continuing the regularity.
  f. One way to make out that argument would be to appeal to the effect of *collateral beliefs*. Suppose it were the case that *every* material inferential-or-incompatibility relation could be infirmed by *some* collateral circumstance. (It would be enough if no content failed to have some

individuation-essential material inferential-or-incompatibility connection of which this were true.) Then unless the regularity *explicitly* incorporated that fact, there would be some possible circumstances in which the two would diverge.

g. [Does the invocation of the effect of collateral belief on this limb of the argument indicate that the two arguments (this one, and the one from holism) really collapse into one, and so forge a connection between Kris's two senses of 'vehicleless'?]

21. The argument from holism for one sense of 'vehiclelessness' of intentional states: Two constraints on the CofI&I of the repeatable *contents* of believings (those contents being believables), which are accordingly also constraints on the individuation of believings:

a. Evans's Generality Constraint concerns the relations between meanings or contents: For Pa to be a possible content of a believing, so must Pb and Qa. 'Cattle', 'cat'—but 'battle'/'bat', 'rattle'/'rat'. But what other things combine with 'cat', 'bat', 'rat'? Sakes.

b. Quine's Meaning/Belief (Language/Theory) Constraint: in order to be able to *mean* that *p*, there will be certain other *beliefs* that one must have (a set, *some* of which one must have).

22. Holism argument for *local vehiclessness*:

a. The idea then is that if *a* can only be the bearer of a content *c* if it stands in relation R to *b*, then really it is not *a* that is the bearer of *c*, but *a* plus *b*. For *a*, by itself, apart from its relation to *b*, does not have the content *c*.

b. If that is right, then the two constraints in (1) mean that only a whole web of beliefs and unbelieved believables (for the Generality Constraint) are going to be the bearers of *all* the contents.

23. But, I think, the argument in (22) is *not* right. What shows that it is not right is *holistic role abstraction*. For HRA shows how holism is compatible with *local vehicles*. [Explain.]

24. The argument from normativity, however, together with the modal separability argument, *does* show that intentional states cannot be *identified with* any non-normatively characterizable states. This is *normative vehiclelessness*. To fill in this argument, we must notice that deviations from regularities, *irregularities*, show that a regularity does *not* obtain, while defications from norms, *errors* or *mistakes*, do *not* show that a norm does not govern the instances. So normative CofI&I are always going to be modally separable from any non-normative ones.

25. How does the modal separability argument bear on the Fodorian "Take care of the syntax and the semantics will take care of itself," view?

a. The sortals that identify and individuate intentional states and episodes seem to underdetermine the issues of identification and individuation, in the sense that they do *not* seem to settle the very modal issues of re-identification that matter for the modal separability argument. Thus I do

not know under what counterfactual circumstances a thinking that p in another possible world would be the *very same* thinking that p that I am engaged in now. At least the content, and most probably the thinker, would have to be the same. But could I have engaged in *that very thinking* at a different time? Only a *slightly* different time? Or a year earlier or later? Could I have engaged in that very thinking—not just had the same *thought*, in the sense of thinking the same thinkable—if my collateral beliefs were different? What if I spoke a different language?

b. The conclusion of that line of thought seems to be that intentional state and episode sortals (note that it seems the same rules do *not* apply to both) are *schematic*, in the way I argued not only 'object' (which is *purely* schematic, or *merely categoreal*), and so 'particular', 'individual', 'entity', 'thing', 'item' and others, but also 'event' and 'process' are.

c. That schematic character leaves room for identification of individual (particular) items picked out by *different*, *less* schematic sortals with items picked out by these more schematic ones. This is an important possibility I had not previously acknowledged. For I had thought that trans-sortal identifications must involve sortals with the *same* criteria of identity and individuation, but possibly different criteria of application. (Note that now I would restrict the latter just to criteria of appropriate *application*, since the CofI&I can be thought of as a subset of the appropriate *consequences* of application.) So there are really *two* ways trans-sortal identifications can be true:

> A) By involving sortals with the *same* criteria of identity and individuation, but possibly different criteria of application; or
> B) By involving sortals one of whose criteria of identity and individuation is suitably *schematic*, and possibly different criteria of application.

d. It is the latter sort of possibility that we might (charitably) see Sellars as envisaging between items in the manifest image and those in the scientific image, supposing that the sortals in the manifest image are suitably schematic.

26. On the same question:
a. But there is still the possibility that the CofI&I of the non-intentional characterization of the vehicles of intentional states and episodes (Is it more plausible that episodes have bearers than that states do? This might be Davidson's view.) already settle it, via the modal separability argument, that trans-sortal *identifications* (as opposed to some other intimate relation, such as being a vehicle not only for the content, but in a different, intimately related sense, for the whole thinking) of tokenings of the sign-design and the thinking are ruled out.

b. We might be able to assess these better, because the sign-design sortals do *not* seem to be schematic in the way the intentional-episode sortals are.

c. Here the question is: Could *this very sign design* occur or have occurred in counterfactual situations in which it was *not* so situated as to *be* a thinking that-p?

d. We have already observed that in asking that question, we must allow *relational* properties of the sign-design to enter into its CofI&I. For if we do not—thinking just of the sign-post as a piece of wood—then the answer is obviously 'Yes', and the modal separability argument tells us that the occurrence of that sign-design cannot *be* (identical to) any thinking.

e. Now semantic externalism tells us that relations to the thing thought about are typically going to be essential to the sign-design being a thinking that-p. But Quinean holism (and the methodology of holistic role abstraction) tells us that its relations to *other* sign designs will *also* be essential. (Horizontal as well as vertical.) And of course *they* also have vertical (externalist) relations essential to *them*.

f. The question is how far out into the world they occupy those relations stretch. If there is nothing in the possible world that *could* not affect the semantic content expressed by a sign-design, by being differently related to it, then it seems the view we are considering must *either* succumb to the modal separability argument *or* degenerate into a *global supervenience* claim, since we will be obliged to say that *that very sign-design* cannot occur in *any other* possible world. (Compare the discussion of Davidson in (3) above.)

g. It is methodologically crucial to this line of thought that the relations we are thinking of the non-intentionally but relationally characterized sign-designs as standing in themselves be characterizable in non-intentional, but quite possible modal, terms. Otherwise the whole program of taking care of the syntax and letting the semantics take care of itself won't fly.

27. So it looks as though we have a choice:
a. Identify the sign-designs that are to be the vehicles of intentional content, and when they are such vehicles, *identified with* intentional episodes (states raise further difficulties), in terms that do *not* individuate them by their relations (even non-intentionally specified) to each other (horizontally) and to features of the world (vertically), in which case the modal separability argument tells us that they (their occurrence) cannot be *identical to* any intentionally characterized episodes—since so individuated, *those very same sign-designs* clearly *could* occur *without* expressing the intentional senses in question.

    or

b. Identify and individuate the sign-designs in terms of their non-intentionally characterized relations to each other (horizontally) and to the world (vertically), thought of (very controversial possibility) as *sufficient*, via holistic role abstraction, to confer intentional content, in which case there is a realistic possibility (modulo the "very controversial" claim mentioned earlier) that they can be identified with intentional episodes such as thinkings, *but* we are embarked on a slippery slope that will lead us to limit the occurrence of such sign-designs to the one world in which they actually occurred. This will block the modal separability argument, but at the cost of our having nothing more than a *global supervenience*

claim about the relation of intentional characterizations to the non-intentional ones from which they are holistic-role abstracted.

28. The claim that there can in principle be non-intentionally characterizable relations among sign-designs (horizontally), and between them and items in the non-sign-design world (vertically), that together are sufficient to isomorphically track the corresponding intentional relations (inferential and referential), so that establishing and following out those relations ("taking care of the syntax") will suffice for the *appearance* of sensitivity to meanings or intentional contents ("the semantics will take care of itself") is itself a substantive and controversial one. (I argued against it in (20) above.)

  a. It is *not*—or at least, it need not be construed as—an *identity* claim concerning ontological items called "intentional contents," or "meanings." For it can be construed as denying that there are such things. What there *really* is is what is *causally efficacious*, and that, according to the methodological motivations of this approach, can always be specified in *non*-intentional, *non*-semantic vocabulary.

  b. On this approach, semantic-intentional vocabulary is a sort of *façon de parler* that is not to be taken seriously ontologically. Semantic-intentional vocabulary can be applied wherever a suitable isomorphism to a suitable set of holistic relations among atomistically specifiable sign-designs obtains. But there is nothing that it *refers to*.

  c. This way of thinking about the formalist-isomorphism-epiphenomenalist strategy for resolving the paradox of mechanical rationality (thought of as a genus with many more determinate species) fits well with Davidson's measurement model or metaphor. When I assign the number 5 to one bag of sand, and 10 to another, I am entitled to do so by various relations those bags stand in to each other, and to other things (including scales). I need not think of the phrase "5 pounds" as referring to anything in order to understand the *use* of this phrase. And we can see this by realizing that we could assign *different* numbers if we used kilograms. And it would make no sense to ask which was *correct*, in the sense of which was *really* there.

  d. But this Davidsonian measuring-with-propositions model seems to differ in just this regard from the holistic role abstraction model. For that *does* give us something referred to by the 'that-p' clauses, namely *equivalence classes* of (possible) atomistically individuated sign-designs.

29.

  a. I claimed last week that 'cause' is also a schematic sortal. To answer the question "Is *a* the same cause as *b*" we need to *supply* supplemental sortals for *a* and *b* (the *same* one for both—or at least ones that share their CofI&I). Here 'event' just puts off the question, since it, too, is a schematic sortal.

  b. Trans-sortal identities of the "water = H_2O" and "lightning = atmospheric electrical discharge" kind—the philosophical paradigms of theoretical reduction—can be understood as making implicit reference to the

schematic sortal 'cause', hence as falling under trans-sortal identity kind (1-c-ii).

c. Note further that these are construed as *necessary* identities, not as *contingent* ones. It is the latter to which the modal separability argument is principally addressed.

d. And they are *type-type* identities, not token-token identities.