PULP NATURALISM

Il Cannocchiale, Rivista di Studi Filosofici, **2** [special issue on Philosophy of Mind and Cognitive Science], 1997: 185-195.

Josefa Toribio
Department of Philosophy
Washington University in St. Louis

Campus Box 1073
One Brookings Drive
St. Louis, Mo., 63130

e-mail: pepa@twinearth.wustl.edu

1. Introduction. The Stand.

There is a compelling idea in the air. Both contemporary philosophers of mind and philosophers of language are engaged in developing theories of (mental or linguistic) content that are **naturalistic**. The stand has been taken: semantic properties are not part of the primitive ontological furniture of the world. If we want to vindicate those properties as real, we will have to show that it is possible to unpack them into some other –primitive– set of properties. It is taken for granted that there is no alternative way of avoiding circularity in explaining the semantic properties of mental or linguistic representations. The following quote from Fodor's Psychosemantics is probably the locus classicus for this trend:

> I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of spin, charm, and charge will perhaps appear upon their list. But aboutness surely won't; intentionality simply doesn't go that deep. It's hard to see, in the face of this consideration, how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. If the semantic and the intentional are real properties of things, it must be in virtue of their identity with (or maybe of their supervenience on?) properties that are themselves neither intentional nor semantic. If aboutness is real, it must be really something else.
>
> (Fodor, 1987, p. 97)

The naturalistic project seems thus to be deeply rooted in the search for a non-circular, explanatory account of intentional categories. Although, in principle, one should not take naturalism in regard to some realm as committing one to any sort of reductive explanation of that realm, in the context of contemporary philosophy of mind and philosophy of language, naturalism and reductionism como easily upon a meeting point. The reason is that explanatory, non-circular answers are taken to be reductionist. If this is correct, we would have to conclude that naturalism entails reductionism (regarding a theory of content) or, even more strongly, that naturalism is constitutively dependent upon the defense of a reductionist thesis. In this case, we would have to admit that there cannot be a non-reductionist naturalistic theory of content.

The main aim of this paper is to cast some light on the issue of what naturalism means regarding a theory of content. It is a matter of a some urgency to find out exactly what it takes for a theory of content to be naturalistic. Until we do so, we cannot properly evaluate the existing theories (or indeed, develop our own). I will center my discussion of this topic on the question of whether naturalistic theories of content ought to be reductionist theories. The claim I want to defend is that, despite the current trend regarding the relations between naturalism and reductionism, the former is not constitutively dependent upon the latter. I will argue that, in fact, the strong requirements of a reductionist thesis are the very reason the project of naturalization seems to be doomed. However, I will try to show that once we have weakened those requirements, non-reductionist answers are certainly acceptable, in the sense of being informative, non-circular, and above all, explanatory. I propose a safe[r] characterization of naturalism that seems to better fit our theoretical views about content, and to turn the whole project into a more promising enterprise.

2. Reductionist Rhapsody.

The question I would like to address is whether a reductionist thesis can solve the problems which those engaged in the naturalization problem want to solve. In order to answer that question, we need to understand first of all what kind of reductionism is at issue here.

A very strong version of reductionism can be construed as follows. We say that a language S_2 is reducible to a language S_1 when we can translate all statements in S_2 into statements in S_1. In order to explain what such translation may consist in, let's imagine a simple case in which the constitutive components of S_1 and S_2 –statements, predicates and proper names– can be arranged according to the same syntactic rules, i.e., that S_1 and S_2, are different only in regard to their vocabulary.

We can then establish some kind of correspondence between expressions in S_1 and expressions in S_2 by using a function ø that assigns, to any basic term A of S_2, a term of the same syntactic category ø(A) in S_1. There is a translation from S_2 into S_1 if and only if, for any basic term A in S_2, A –under the interpretation of the language S_2– is synonymous with ø(A) –under the interpretation of the

language S_1– (Cfr. Kutschera, F, 1982, pp. 165-179). Dummett's characterization of reductionism follows the same line:

> Reductionism, properly so called, is the thesis that there exists a translation of statements of the given class into those of some other class, which I shall call the reductive class. This translation is proposed, not merely as preserving truth-value, but as part of an account of the meanings of statements of the given class.
>
> (Dummett, 1982, p. 66)

Now, the main problem with this strong version of reductionism is that it doesn't fit the picture that naturalist theorists like Fodor have in mind. If the type of reductionism involved in the project of naturalizing semantics were of this full-fledged kind, we should be looking for a translation scheme (a translation function ø) from S_2 to S_1 such that for any basic term A in S_2 (under the intentional interpretation of the language S_2) there is a term ø(A) in S_1 (under the physicalist interpretation of S_1) such that ø(A) is synonymous with A. However, it is clear that no one out there is looking for translation schemes as part of naturalization.

Maybe then it is just an unfortunate choice of terminology. Maybe what naturalization is all about has very little to do with reductionism –properly understood– and much more with the defense of what Dummett calls a reductive thesis (henceforth, to make it less confusing, weak-reductive thesis), what McGinn calls a reduction of truth (cfr. McGinn, 1982, pp. 120-122) and what almost everyone calls a supervenience thesis. The task of a weak reductive thesis (or a reduction of truth or a supervenience thesis) is to specify what makes a sentence of the given class (S_2) true, when it is true:

> A [weak] reductive thesis ... claims only that no statement of the given class can be true unless some suitable statement or statements of the reductive class are true, and, conversely, that the truth of those statements of the reductive class guarantees the truth of the corresponding statement of the given class.
>
> (Dummett, M., 1976, p. 94)

If "reductionism" is then not the proper term to refer to the characteristic strategy of the naturalistic enterprise, I should reformulate my question in the weak-reductive sense that seems to describe it better. The question would thus be whether naturalistic theories are constitutively dependent upon the defense of a weak-reductive thesis.

3. The Remains of Reductionism.

A weak-reductive thesis involves a reduction of truth of the statements of a given class to the truth of the statements of some other (reductive) class. Ernest Nagel's model of reduction in Chapter 11 of The Structure of Science (Nagel, 1961) fits this weak-reductive picture. Nagel's terms of reduction are not those of statements or classes of statements. Following the positivist spirit of his time, the account he provides applies to reduction of one science by another –and ultimately of all sciences by physics. What Nagel's reduction involves is thus the explanation of the laws of a higher-level science by the laws of a lower-level science through the deductive application of bridge laws.

It might be argued that Nagel's model of reduction in terms of reduction of laws seems prima facie to be yet again too strong for the purposes of naturalizing semantics. But, regardless of the nomological character of the statements in the sciences at issue, there is something in this model that is central for the weak-reductionist: explicability. Even a weak-reductionist wants a science or a given class of statements be explicable (and causally explicable) in terms of some other science or class of statements, and again ultimately explicable by physics (See Brooks, 1994).

I shall return to this point shortly, but let me say first that if naturalistic theories of content are constitutively dependent upon the defense of this kind of weak-reductive thesis, those theories will have to provide, for any statement A in the class of intentional statements, some other statement (or some family of statements) A' in the class of physical statements such that, for A to be true, it is at least sufficient that A' be true (cfr. Dummett, 1976, p. 94). However, even under this new formulation, the project of naturalization seems to face a few important hurdles.

The one I would like to concentrate on has two versions. One is the impossibility of specifying (even) sufficient properties (and, a fortiori, necessary

and sufficient properties) that would constitute the parameters of individuation for the intentional content of a token thought (or token utterance) at the physical level. It has been argued that this impossibility is, almost by definition, a limitation of conceptual analysis. Work in cognitive science on concepts and categories (cfr. Rosch, 1973, 1975, 1978; Smith and Medin, 1981) suggests that intentional concepts simply are not constituted in terms of necessary and sufficient conditions. This line of argument has been exploited by, e.g., Stich and Laurence (Stich and Laurence, 1994) and Tye (Tye, 1992). I will not dwell on it here, although it is interesting to note that naturalist-inclined theorists seem to be content with the specification of merely sufficient conditions. So, the fact that no necessary and sufficient conditions can be found would not be a real problem for them.

The second version of the hurdle, the one I would like to comment on here, has been beautifully formulated by Terence Horgan:

> Although a physical supervenience base might always exist for any manifestation of aboutness, in general any adequate non-intentional, non-semantic characterization of the supervenience base might be enormously baroque and complex. Perhaps, for instance, the supervenience base for the intentional content of a token thought (or token utterance, or token inscription) generally involves a good-size chunk of space-time extending well beyond the cognizer's own body and well beyond the time at which the token thought occurs; perhaps it involves a rather gargantuan number of physico-chemical goings-on within that extended spatio-temporal region; and perhaps there isn't any simple way to describe, in non-intentional and non-semantic vocabulary, all the relevant aspects of this hugely complex supervenience base.
>
> (Horgan, 1994, p. 309)

Using our favorite terminology: the problem is that even the defense of a weak-reductive thesis regarding intentional properties of mental and linguistic representations is blocked by the fact it is extremely difficult (read impossible) to specify the statements in the reductive class whose truth will guarantee the truth of any statement selected from the given class. That being the case, we can't expect to be able to pick out precisely those entities/properties that would count as the supervenience base for any token thought or token utterance.

Now, let me refine this position a bit. It could be argued that the advocate of a weak-reductive thesis doesn't necessarily have to face this problem, that one might e.g. give a functional or causal definition –not in the language of physics– which would show how physics could entail intentional properties all the same. All we need, in other words, is a functional partition of those categories that play a role in semantics. Those categories will be thus characterized in terms of their functional/causal properties without having to spell out the structural/physical properties underlying them.

What I would like to stress here is that, although within this kind of weak-functional-reductive thesis we give up the need for a specification of the properties that constitute the supervenience base of token thoughts or token utterances, the explanatory spirit of the full-fledged reductionism remains intact. What is expected from a naturalization proposal of this kind is an array of causal explanations. Even if the reduction to physics is avoided, the explanatory model of physics is maintained. The explanations within the weak-reductive model are still very much like the explanations in physics and that poses important problems for the naturalization project (see below).

We don't have to conclude, however, that the project of naturalizing the intentional is doomed. I want to suggest that the logical conclusion is that all naturalization requires is that one be able to explain how intentional properties arose from non-intentional ones. It doesn't require anything stronger, such as reducing intentional properties to functionally-defined causally efficacious properties. Under this non-reductive interpretation of the constraints for naturalism, we need not affirm that all facts which can be described and explained with the help of theories involving intentional language, can be described and explained in a purely physicalist language.

4. Safe[r] Passage.

One way of unpacking the notion of explanation invoked above is the following. We can say that a non-reductive, naturalistic account of intentional properties is one that is consistent with current empirical theories regarding cognition, i.e., one that is compatible with the knowledge that different empirical theories (biology and neuroscience, mainly) have to offer. An account, in order to be naturalistic, should not conflict with what we know about how the brain works and/or with what we know about the scientific role of biological functions. But this is clearly too weak. After all, someone might argue, divine intervention probably is consistent, that is, it doesn't conflict with current empirical theories regarding cognition and yet it would be quite surprising for a naturalized account of any kind of process to include supernatural powers within its theoretical apparatus. The claim ought to be stronger. What we need is a type of explanation which shows how brains and biological species could do what our account of content requires. The claim I want to defend is that what makes a theory of content a naturalized theory of content is that the notions invoked in the theory capture structural features of the cognizer's intentionality which make it behave in ways that fall under a scientific explanation.

We have to be careful though about the notion of scientific explanation. A scientific explanation doesn't necessarily involve physical descriptions. Even if it does, it is simply not true that physics doesn't include quite abstract theoretical entities as part of its most basic explanations. In that sense, the version of naturalism present in Fodor's quote, as representative of a widespread contemporary trend, is based on the misconception that physical descriptions include only theory-independent, real objects and properties. It is interesting to notice in connection with this point that what seem to be powerful counterexamples to the idea that good explanations are always causal explanations are taken frequently from quantum mechanics (see Salmon, 1984). But even if those cases could not be regarded as successful counterexamples (I must admit that my knowledge of quantum mechanics is too naïve to be a good critic here), there is a more important misconception about scientific explanations that I'd like to stress. This misconception is to take causal explanations, i.e., the

kind of explanation commonly associated with physical explanations, as the only model of explanation. Yet there seem to be good reasons to believe that causation is just a particular kind –very important but not unique– of determinative relation. I think G. MacDonald is right when he claims:

> Some of the problems which have been thought to plague attempts to naturalize the mental ... arise out of a prejudice which restricts the proper form of scientific explanation to causal explanation.
>
> (G. MacDonald, 1992, pp. 242-243)

Think of what Kim calls Cambridge dependency (cfr. Kim, 1974). Socrates dies and, in virtue of this, Xantippe becomes a widow. Socrates' death and Xantippe's new marital status are not the same event. The first difficulty with a causal account of the relation between these two events is that they are simultaneous. Also, the relation doesn't seem to instantiate any nomic regularity, that is, there doesn't seem to be any empirical law that could support a causal relation between Socrates' death and the widowing of Xantippe. What we have here is some kind of logical entailment, not a physicalist or causal connection: "Thus, one might say that the proposition that the death of Socrates occurred at t, taken in conjunction with the standing condition that Socrates was the husband of Xantippe at t, entails the proposition that the onset of Xantippe's widowhood occurred at t" (Kim, 1974, p. 43).

I said "some kind of logical entailment" because the two events are not symmetrically related, as they would be if we took the logical entailment at face value, i.e., if we took the relation to be that of a biconditional. Xantippe's widowhood depends on Socrates' death, but not the other way around. This asymmetric dependency can be better appreciated if we realize that while it seems intuitively true that

> If Socrates had not died at t, Xantippe would not have become a widow at t,

the relation expressed by

> If Xantippe had not become a widow at t, Socrates would not have died at t

seems to reflect an alteration in Socrates' marital status instead of the relation we want to capture (See Kim, 1974, p. 43). That relation –the one between Socrates' death and the widowing of Xantippe– is an explanatory one, but not a causal-explanatory relation. Although there is no causal connection between Socrates' death and Xantippe's widowhood, the former event explains the latter. Cases like these clearly provide good examples of non-causal, good explanations.

It is this kind of determinative relation that I had in mind when I claimed that a naturalized theory of content should treat intentional properties in a way that makes them liable to follow patterns that fall under scientific explanation. But scientific explanation doesn't necessarily have to be causal explanation. We don't have to face thereby the problems involving the correct individuation of properties that will instantiate some nomic regularity, as most of the time there will not be empirical laws supporting that kind of determinative relation.

The main difference between a weak-reductive thesis and my position lies in this point about causal explanations. While a weak-reductivist is still looking for causal explanations in order to achieve a naturalized semantics –even if her approach is a functional one–, someone engaged in my safe[r] naturalistic enterprise would be content with Cambridge dependencies à la Kim between intentional and non-intentional properties. The consistency between an intentional and a non-intentional story that is required by any naturalistic project in semantics ought to be cashed out in terms of the theoretical criteria that guide us in dividing up the accounts that clearly are not in consonance with current empirical research about cognition, from those that are consonant with that empirical orientation.

Under this new specification, there does not seem to be any reason to think that it is in fact the case that the intentional cannot be naturalized. If there were, i.e., if we were to fail to achieve a naturalistic account of the intentional in the safe[r] sense of naturalism defended here, we would indeed bring forth the demons described by Fodor. It would be tantamount to giving up a place for the mind in the natural world, the only world we have got. So we had better keep working on it.

Acknowledgments

I would like to thank Andy Clark and David Chalmers for useful comments on an earlier version of this paper.

References

Brooks, D. H. M. (1994), 'How to Perform a Reduction', Philosophy and Phenomenological Research **54** (4), pp. 803-814.

Dummett, Michael (1976), 'What is a Theory of Meaning? (II)', in G. Evans and J. McDowell, eds., Truth and Meaning. Essays in Semantics, Oxford: Clarendon Press, pp. 67-137.

Dummett, Michael (1982), 'Realism', Synthese **52** (1), pp. 55-112.

Fodor, Jerry (1987), Psychosemantics, Cambridge, MA: MIT Press.

Horgan, Terence (1994), 'Computation and Mental Representation', in S. Stich and T. A. Warfield, eds., Mental Representation. A Reader, Oxford: Basil Blackwell, pp. 302-311.

Kim, Jaegwon. (1974), 'Noncausal Connections', Noûs **8**, pp. 41-52.

Kutschera, F. von (1982), Grundfragen der Erkenntnistheorie, Berlin, New York: Walter de Gruyter.

MacDonald, Graham (1992), 'The Nature of Naturalism', Proceedings of the Aristotelian Society, Suppl. vol. **66**, pp. 225-244.

McGinn, Colin (1982), 'Realist Semantics and Content-Ascription', Synthese **52**, pp. 113-134.

Nagel, Ernest (1961), The Structure of Science, London: Routledge & Kegan Paul.

Rosch, Eleanor. (1973), 'On the Internal Structure of Perceptual and Semantic Categories', in T. E. Moore, ed., Cognitive Development and the Acquisition of Language, New York: Academic Press.

Rosch, Eleanor (1975), 'Cognitive Representation of Semantic Categories', Journal of Experimental Psychology: General **104**, pp. 192-233.

Rosch, Eleanor (1978), 'Principles of Categorization', in E. Rosch and B. Lloyd, eds., Cognition and Categorization, Hillsdale, NJ: Erlbaum.

Salmon, W. (1984), Scientific Explanation and the Causal Structure of the World, Princeton: Princeton University Press.

Smith, E. and Medin, D. (1981), Categories and Concepts, Cambridge, MA: Harvard University Press.

Stich, Stephen. and Laurence, Stephen (1994), 'Intentionality and Naturalism'. Midwest Studies in Philosophy.

Tye, Michael (1992), 'Naturalism and the Mental', Mind **101**, pp. 421-41.