# NATURALISM AND CAUSAL EXPLANATION

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

Semantic properties are not commonly held to be part of the basic ontological furniture of the world. Consequently, we confront a problem: how to 'naturalize' semantics so as to reveal these properties in their true ontological colors? Dominant naturalistic theories address semantic properties as properties of some other (more primitive, less problematic) kind. The reductionistic flavor is unmistakable. The following quote from Fodor's *Psychosemantics* is probably the contemporary *locus classicus* of this trend. Fodor is commendably unapologetic:

> 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, 97)

Notice the shape of this explanatory project. Intentional properties will count as real in virtue of their identity with, or supervenience on, some set of lower-level physical properties. Fodor thus assumes, in effect (as do many others engaged in naturalization projects for semantics), that the program of naturalization demands a higher-to-lower, top-to-bottom, kind of explanatory strategy.

This paper addresses precisely that assumption, namely, that the non-semantic properties on which semantic properties depend, belong to what are intuitively lower levels of description than the intentional level itself. It also questions the higher-to-lower

explanatory scheme associated with that assumption. My discussion of this topic draws on Robert Brandom's recent work (Brandom 1994) and can be considered an analysis of Brandom's stance and its implications. The discussion should help to explain the general lack of progress in the project of naturalizing content. It should also help show why attempts to eliminate the normative vocabulary employed in specifying the practices that guide the use of a language are unlikely to succeed.

I shall start by displaying the general order of explanation that characterizes typical naturalization projects, showing that even when a full reduction to physics is avoided, some important assumptions inherited from the explanatory model of physics remain. These include the demand for an array of *causal* explanations couched in terms of ultimate properties of the world, and the idea that such non-semantic properties should be *constitutive* (in a narrow or individualistic sense to be explained below) of whatever semantic properties are in question. Extending Brandom's idea that the normativity of content is not reducible to physics, I shall argue that even such residual demands are inappropriate. More positively, I suggest that, despite the deep irreducibility of the normative dimension of content, we need not consider that dimension either primitive or inexplicable. Instead, such normative aspects can be unpacked by invoking a different, lower-to-higher, explanatory scheme in which the *explanans* includes higher level features such as skilled know-how and social frames of action.

## 2.    Explanation and Causal Explanation

Classic positivist models of reduction involved the *explanation* of the laws of a higher-level science by the laws of a lower-level science through the application of bridge laws. The aim was to reduce one science to another —and ultimately all of science to physics (Cf. Nagel 1961). This positivistic model is now perceived to be too strong for the purposes of naturalizing semantics. However, one important aspect of that model still informs most attempts to naturalize semantics: this aspect concerns the *causal* explicability of

higher level properties by lower level ones (Cf. Salmon 1984). The semantic naturalist demands that intentional properties be *causally explicable* in terms of some other class of more basic properties, even if there is no ultimate reduction to physics (Cf. Brooks 1994).

Recall the opening quote from Fodor —taken as representative of a widespread contemporary tendency. The explanatory strategy, in practice, consists in depicting linguistic intentionality as derivative from mental intentionality, and mental intentionality as derivative in turn from more basic cognitive properties. The content of a linguistic expression is derived from some array of cognitive capacities and those cognitive capacities are seen as the result of some rather complex organization of human wetware. This putative chain of dependencies between mental and linguistic content helps us to understand why contemporary philosophy of language and contemporary philosophy of mind have become more and more a single discipline. As linguistic properties are seen as derivative from properties of contentful mental states, the study of content focuses mainly on mental representations. An account of mental content is taken to be essential to providing an account of linguistic content.

Now, supposing that intentional content is nothing but the *meaning* of mental representations and that internal processes are only sensitive to the syntactic structures of mental representations, not their meaning, one of the first problems the semantic naturalist has to face is how to uphold the idea —absolutely plausible from an intuitive point of view— that the intentional content of a thought (or any other intentional state) is *causally* responsible for behavioural (and other) outputs. It quickly seems that if we want to uphold such an idea, and to do so on a good physicalist basis, we will need to postulate the existence of micro-physical mechanisms or structures corresponding to different types of semantic events and subsumable by laws. The semantic event-types are thus considered causally efficacious only insofar as they supervene on micro-physical events —perhaps different ones on different occasions— in such a way that the causal powers of the former are *explained* by the causal powers of the latter.

We need to talk about *types* of events because, although only individual events can be causes, the explanatory character of the regularities captured by a causal law depends on such regularities

being established not between particular events, but between *types* of events. Now, since particular events can be referred to by many different expressions, some of which don't mirror any of the properties that turn them into causes or effects of other events, the criteria for grouping together particular events into events of the same type —events of the type that can appear in a causal law— must only focus on those properties that can be shown to be causally efficacious. To take a non-semantic example, although it might be true that the dinner on Sunday caused my stomach ache, the causal law on which such a truth depends doesn't invoke any relation between events of the type *to have dinner on Sunday* and *stomach ache*, but rather between events of a particular physical type, such as a certain composition of the food and an alteration in the digestive juices. Of course, when one applies the same strategy to semantics, the immediate risk is that of epiphenomenalism. In other words, the risk is that intentional states considered *qua* intentional, (i.e., considered as having a particular meaning), fail to appear causally responsible for our behavior because the semantic properties themselves don't contribute to the causal powers of the states to which they belong. The recent epidemic of papers concerning the causal efficacy of content[1] reflects in part the perceived urgency of the search for a solution to the threat of epiphenomenalism; the urgency, that is, of the search for a physical causal-explanatory basis able to embrace semantic facts as properly explanatory of behavior.

Yet despite this very strong tendency, there are good reasons to believe that causation is just *one* particular kind —important but not unique— of determinative relation. In this vein, G. MacDonald comments that '[s]ome 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' (MacDonald 1992, 242-243). Thus consider what Kim calls *Cambridge dependency* (Cf. Kim 1974). Socrates dies and, in virtue of this, Xantippe becomes a widow. Socrates' death and Xantippe's new marital status are surely not the same event, but what kind of relation holds between them? One 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, i.e., 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, 43).

I said 'some kind of logical entailment' because the two events are not symmetrically related, as they would be if we took the notion of 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 far less commanding of our assent (Kim 1974, 43). The relation we wish to capture —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 seem to constitute examples of good, non-causal, explanation. And what they show is that, once we look closely at the very notion of explanation, we see that 'explicability' is not necessarily to be equated with '*causal* explicability'. Of course, this is not to say that we should abandon the notion of causal explanation. The point is only that other forms of explanation —that establish determinative but non-causal

relations between events— offer perfectly acceptable ways to render certain target properties non-mysterious.

### 3.    Institution *versus* Constitution.

It could be argued that Cambridge dependencies á la Kim are good explanations only in an analytic, trivial sense. It's only because it's part of the meaning of 'widow' that one's husband has died that we consider Socrates' death as a good explanation of Xantippe's widowhood. This is an interesting objection, rooted, indeed, in the very fact that non-causal determinative relations of this kind do not figure in standard explanatory frameworks. But two points should help establish that the sense of triviality is only apparent. Firstly, what we understand by the term 'widow' depends very much on a range of social customs and institutions. It wouldn't make much sense (or, at least, it would make a *different* sense) to talk about widowhood in a society with no institution of marriage or in a society where polygamy was imposed on women. Under those circumstances, Kim's example wouldn't be a good explanation because there would be no determinative relation between Socrates' death and Xantippe's widowhood. Cambridge dependencies thus introduce a reversal of the usual order found in contemporary explanatory schemes. Not only do we give up the equivalence between 'explanation' and 'causal explanation', we also give up the higher-to-lower character of the explanation involved. Here, the *explanans* does not include properties from a lower level of description. On the contrary, it includes properties that belong to a much higher —in this case, social / institutional— level. If we look back at Fodor's quote, we realize that all that is required to make semantic properties non-mysterious is that they be accounted for in some non-intentional, non-semantic terms. The additional idea that those non-intentional, non-semantic properties must belong to a lower level is an optional extra. We must, of course, avoid circularity in the explanation of intentional affairs. But it is nowhere written that circularity can't be avoided by some kind of upwards-looking relation instead.

The second, and more important, point is that the rupture of the higher-to-lower explanatory strategy calls into question one of the core metaphysical notions in the framework of scientific explanation, the metaphysical notion of *constitution*. Let me dwell on this a little. The notion of constitution I have in mind is the one at

work in the thesis that what makes a thing into a thing of a certain kind is the relation between that thing and the stuff it is made of. Although there are other, more contextually oriented, ways of accounting for the exact nature of a particular thing (see below), *constitution*, understood in this narrow and *individualistic* sense, seems to be at the root of the model of explanation that informs most contemporary naturalistic proposals. It is this notion of constitution that demands that we cash out the connections between the internal properties of an entity and its surroundings in *causal* terms. Roughly, this is how the theses fit together. A particular piece of metal is constituted as e.g. a piece of gold if it is made of the right stuff, if its composition has the atomic number 79. What it is for a piece of metal to be gold is then fully *explained* by appealing to the properties of its internal structure. But it is also its internal structure that is important in accounting for whatever causal powers that piece of gold might have.

This notion of constitution ought to be distinguished from another kind of relation, also sometimes called a constitutive one, but which is more logical than metaphysical. The *logical* notion of constitution amounts to something like a set of necessary and sufficient conditions that something has to meet for being a thing of a certain kind. In current debates, the phrase 'being constitutive of' —as opposed to 'constituting'— tends to refer to this logical cousin of the metaphysical notion. Thus even though it is constitutive of the number 2 to be the successor of the number 1 (in the logical sense of constitution), 2 is not constituted (in the metaphysical sense) by being a successor of 1. Most naturalistic approaches such as Functional Semantics (Block 1986), Interpretational Semantics (Cummins 1989), Information Based Semantics (Israel and Perry 1990; Dretske 1988), Asymmetric Dependence Theory (Fodor 1990), Teleological Theories (Millikan 1984; Neander 1991) all aim to discover some set of lower level, non-semantic, properties in terms of which semantic properties are metaphysically *constituted*, i.e. those properties that make a given representation into the representation it is and that will explain why it is that representation (and hence why it has the semantic properties it has) rather than some other. Clearly the aforementioned attempts at naturalization are not looking for constituting properties in the

*logical* sense. The search is not for something like necessary and sufficient conditions. The search is rather for those non-semantic properties that play a (metaphysical) *constitutive* role in the instantiatiion of semantic properties. Underlying the model of explanation at work in contemporary naturalistic proposals there is thus not just the general physicalist bias mentioned earlier, but also the additional idea that what makes a thing into a thing of a certain kind is the relation between that thing and the stuff of which it is made.

To see this notion of constitution in action and to see how it invites the observed emphasis on causal relations, think of e.g. the internalism / externalism debate in philosophy of language. The internalist often agrees that e.g. the individuation of the contentful states of cognitive systems is somehow relative to the properties of the world that they inhabit. But for her, this is just to say that the world is what determines the agent's cognitive states (usually via sensory inputs), and not that external features play any role in *constituting* such states. In fact, one of the most characteristic features of internalism is precisely the metaphysically individualistic claim that only local features of a given mental representation are relevant to its constitution. The justification of the internalist view thus involves the idea that the causal powers of any event are completely determined by its local physical features. As a result, only content that is individuated in terms of a system's *intrinsic* properties is deemed adequate for a scientific explanation of behavior.

Now, as I said above, there is a more contextually oriented alternative to this way of accounting for the exact nature of a particular thing. John Haugeland and Tim van Gelder are among those who have most recently explored this possibility (Cf. Haugeland 1993 and van Gelder, 1993), referring to it as the 'holistic' alternative:

> Roughly, holists see the fundamental nature of things as depending only upon some larger whole to which they belong ... metaphysical holists see constitution as solely a matter of context; an individual entity is what it is in virtue of some larger structure or pattern into which it fits ... If an entity is

constituted as an A holistically, then to understand that entity as an A you have to understand the relevant larger whole and how As fit into it. To render an A intelligible as an A is to articulate its place in the larger whole.

<div align="right">(van Gelder 1993, 67)</div>

To avoid confusion between the individualistic and the holistic notions of constitution, I propose to label the latter notion 'institution'. Institution is thus a context-oriented way of understanding what makes a particular entity into the entity it is. To say of something that it is instituted by such-and-such properties is to claim that its fundamental nature in best accounted for in terms of the properties of some specific context in which the entity is manifest, and thus to understand the entity as an entity of that particular kind involves an appreciation of the properties of that context. John Haugeland gives the familiar example of the ontology of chess pieces (Haugeland 1993, 4). Suppose we ask what makes something a queen in chess? Obviously queens are made of some material or another but this individualistic notion of constitution will not help us understand the fundamental nature of a queen. What makes something a queen can only be fully understood by focusing on the role that the piece plays in the larger game of chess and hence by focusing on contextual properties such as the rules of chess[2].

Constitution and institution are not, of course, exclusive ontological relations. In most cases, that something is an entity of a particular kind results partially from the stuff it is made of and partially from the (environmental, social, legal, political, etc.) context in which the entity is located. However, there are countless entities for which the only way of getting at what they really *are,* and the only way of explaining why they behave how they do, is to look at the contexts that *institute* them, not the stuff they are made of. This is obviously true in the case of social institutions. Their nature, character and function are completely dependent on the socio-political context and cultural practices of a country or a community. Even though these are obviously instantiated by organizations of matter of some kind or another, the stuff they are made of doesn't make them the entities they are. And it certainly

doesn't help to explain why they behave in one way rather than another.

By making institution the metaphysically central notion in our explanatory accounts, we change the higher-to-lower explanatory strategy common to most naturalistic approaches mentioned above. The sense of triviality that our objector found in the Socrates' example seems to disappear when we thus change the character of the explanatory framework from the constitutive to the institutional. Socrates' death is indeed a good explanation of Xantippe's new marital status, in part because of the meaning of 'widow'. The important point is that this meaning is not itself constituted by any internal physical properties of the agent who has thoughts about widows or who utters sentences in which the the word 'widow' appears. What determines that the term has the meaning it has is a largerly institutional matter, dependent mainly (although not exclusively) upon the social practices and activities of a community. That doesn't render the determinative relation that holds between Socrates' death and Xantippe's widowhood explanatorily empty. Rather, it changes the character of the explanatory framework itself.

## 4.    Lower-To-Higher Models of Explanation

Unfortunately the claim that meaning is a purely institutional matter is to some extent just trivially true. Languages, *qua* systems of interpersonal communication, are obviously social phenomena. However, the claim and the extent to which the social dimension is of philosophical interest can be made more precise. Wittgenstein, Wittgensteinian and, in general, post-Hegelian philosophers have made powerful efforts to clarify the philosophical implications of such a 'trivial' claim. Among the latter, a recent example is Robert Brandom's analysis of linguistic content (Brandom 1994). This is one of the very few exceptions to the standard pattern of physicalist explanation criticized above, as is Adrian Cussins' work on the construction of meaning (Cussins unpublished). Both projects defy easy rehearsal. But they stand out as existence proofs of the potential richness of the alternative explanatory strategy. The

present paper is, in a sense, an exploration of the potential relevance of such work for the more general topic of naturalistic explanation in semantics.

Brandom's project is to develop a pragmatic account of the content of the expressions of a language framed in terms of a notion of commitment. Commitments are characterized by a kind of deontic status and deontic statuses are *instituted* (i.e., holistically constituted) by social practices. The metaphysical notion of institution (the notion of holistic constitution) is thus at the core of Brandom's pragmatic account of content: "our activity *institutes* norms, *imposes* normative significances on a natural world that is intrinsically without significance for the guidance or assessment of action' (Brandom 1994, 48). Adrian Cussins, in the same vein, claims:

> Cognizers participate in the maintenance and construction of normative activity systems which are larger, more complicated, more capable, and perhaps more meaningful than they themselves are. These systems —loci of meaning and value— include institutions, historically established cultures, social customs and norms, built environments, friendships, technologies, languages, arts and artefacts and facts.
>
> (Cussins unpublished, 4)

Cussins' project is an expression of what he calls 'naturalized transcendentalism'. But what is most interesting for our purposes is, again, the fact that institution lies at the core of the posited relations between representation and reference. Material, social and historical contingencies institute the complex network of normative activity systems that make human cognizers what they are.

To say that meaning is normative is to say that to know the meaning of an expression is to know how to use it correctly —even if that knowledge is not explicit. Any account of meaning should thus provide a set of constraints to which correct uses must conform. Satisfactory accounts of meaning should establish *how* the speakers' knowledge of the meaning of the expressions of a language enables them to use those expressions correctly. The key move in both Brandom's and Cussins' proposals is to make

normativity a central and irreducible feature of meaning. The goal of accounting for semantic properties in a naturalistic way thus becomes the goal of preserving (not discharging) that essential feature. As a result, the standard explanatory strategy is altered in many respects. First, no reductionist assumptions motivate this way of accounting for semantic properties. The normativity of linguistic content is the starting point and is treated as ineliminable. Second, although the normative dimension of language is not eliminable, it can be *explained*, so it is not treated as some kind of bare primitive. Third, the notion of explanation is not physicalistically biased; it is not framed in individualistic constitutive terms. Overall, then:

> ... norms are understood as *instituted* by social-practical activity ... [and] ... rendered less mysterious ... by explaining exactly what is expressed by normative vocabulary ... This is an explication of explicitly normative conceptual contents in terms of implicitly normative practices, rather than a reduction of normative terms to nonnormative ones.
>
> <div align="right">(Brandom 1994, xiii-xiv)</div>

Both Brandom's and Cussins' projects involve myriad layers of complexity. Nonetheless the message is clear enough: there is indeed an alternative explanatory scheme to the prevalent higher-to-lower one present in standard naturalization approaches. The point is not that we should abandon the standard, physicalistically biased, model of explanation altogether. It is rather that other, equally respectable explanatory strategies may be especially relevant to the treatment of meaning-based phenomena.

## 5.    Naturalism and Normativity

So far, I have made explicit a general model of explanation that informs many naturalization projects in semantics. I have pointed out two important underlying physicalist assumptions governing that model: the demand for a causal, higher-to-lower explanatory strategy and the key role played by the metaphysical notion of individualistic constitution. I have also noted an alternative

strategy in which explanatory relations are grounded in the holistically characterized, socio-pragmatic notion of institution.

I would like to take now a further step and suggest that such non-causal, institution-based, approaches to explanation should be generally preferred to causal/individualistically constitutive models in the special context of naturalizing semantics. To make the case, I shall focus on one of the main problems that faces standard naturalization projects: the problem of demarcating cases of malfunction from cases of misrepresentation, and of distinguishing the latter from correct representation. This has been a recurrent theme, especially in naturalistic proposals of the informational-teleological kind. Attempts to solve this problem have turned the literature into a kind of intricate scholastic debate illuminated by a great profusion of baroque counterexamples. But there have been very few attempts to question the basic framework that shapes the problem itself.

How can a mental state be said to misrepresent without presupposing any semantic notions? When we say of an artifact such as e.g. a thermostat that it misrepresents the temperature of the room, we don't ascribe to it any original representational powers. What we are saying is just that there is a malfunction in the artifact. As Dretske notes (Dretske 1986), artifacts don't have 'original representation'. If we are searching for original representation, one place to look is towards evolved structures whose adaptive role is that of gathering the information necessary for the creature's survival (it is also essential, for Dretske, that those structures allow the organism to be capable of associative learning). We thus first define a relation that beliefs bear to properties that are sometimes instantiated in the system's environment as the relation 'has as its content that'. The instantiation of a given property then explains why the production of a belief token in certain conditions helps the system perform its proper function. Instantiations of a property causally affect the system and are either developmentally (as in Dretske 1986) or evolutionarily (as in Millikan 1984) relevant with respect to the proliferation of the system. Once all this is in place, false beliefs are explained in terms of the direct malfunction of the evolved

structures or the alteration of the usual conditions in the environment (Cf. Millikan 1984)[3].

Unfortunately, most situations in which an instantiated property counts as the cause of a given belief token are also situations in which other properties, co-instantiated with the first, are also possible causes of that belief token. This is, roughly, the so-called 'disjunction problem' (Cf. e.g. Dretske 1986 and Fodor 1990). The disjunction problem arises whenever a theory can't distinguish between a true token of a symbol that means something disjunctive, and a false tokening of a symbol that means something non-disjunctive. According to Fodor, the problem is so critical that the success of the whole project of naturalizing semantics depends on finding a solution within a naturalistic framework. Fodor gives the following gloss on one fairly typical, but ill-fated, kind of response to this problem:

> [perhaps] what determines the identity of the concept the student has learned is not the actual and counterfactual distribution of his tokenings, … but rather the distribution of actual and counterfactual *punishments and rewards* that prevails in the training situation. So, for example, imagine a student who has been reinforced for positive responses to *apples*, and suppose that no *wax apples* have been encountered. Then what determines that the student has learned the concept APPLE rather than the disjunctive concept APPLE OR WAX APPLE is that, *were he* to respond positively to a wax apple, the teacher (or some other environmental mechanism) *would contrive to punish the response*.
>
> (Fodor 1990, 63)[4]

It is, indeed, not uncommon to find that responses to the disjunction problem from within an informational-teleological framework involve the invocation of such additional normative apparatus. The presence of an actual teacher is, of course, optional. 'Other environmental mechanisms' may do the job. But some externally determined notion of right and wrong response is always invoked —one whose reducibility to properties recognized by physics

is far from obvious. As a result, one basic problem that critics find with informational-teleological proposals is that they reflect an unstable mixture of reductionism (recall Fodor's opening quote) and the reluctant appreciation of the need to invoke some additional, genuinely normative, dimension for the explanation of meaning (see e.g. Smith 1990 and Loewer 1994)[5]. Informational-teleological accounts are thus in trouble because, despite their naturalistic aspirations, they are forced to invoke relations with a very clearly normative character (punishment and reward) —a normative character that cannot help but compromise the project of physicalist reductionism itself.

Such problem, I think, will recur *whenever* we try to naturalize semantics following a higher-to-lower explanatory strategy. For the individualistic notion of constitution upon which standard naturalization projects are built is so determinedly physicalist that it cannot, ultimately, make contact with the normative features it is trying to explain. Standard projects are thus forced to 'cheat' by impregnating the merely physical with normative notions that don't belong there. By treating normativity as ineliminable from semantics and by understanding normativity as instituted by practical activities in a social world, we avoid such a double jeopardy. The committed naturalist, if I am right, should thus whole-heartedly embrace an alternative explanatory strategy, one which locates the normative dimension of intentionality in its natural home: the conventions and constraining practices of a society of intelligent beings.

## Footnotes

1.     See e.g. Fodor 1989, Jackson 1996, Jackson and Pettit 1990, Lepore and Loewer 1987, Segal and Sober 1991.

2.     Of course, similar considerations have been invoked in support of functionalism, a paradigmatically internalist position in Artificial Intelligence. The difference is that, for the functionalist, the role involves only permissible moves in a inner economy, whereas, for the institutionalist, the roles are defined in terms of wider, usually socio-pragmatic, settings.

3.     There are, of course, important differences between Dretske's informational and Millikan's teleological accounts. Dretske is mainly concerned with how causes local to an organism give rise to the content of that organism's particular thoughts. Millikan, by contrast, stresses evolutionary history. These differences, however, need not concern us here.

4.     Fodor's gloss claims to reflect Dretske's favored response to the disjunction problem. Our interest, however, lies not in the accuracy of any such attribution, but in the general shape of a fairly wide variety of naturalist responses —a shape nicely captured by Fodor's description.

5.     Interestingly, Fodor's own causal proposal has been criticized for exactly the same reasons, namely, that no causal condition can explain reference except in conjunction with further intentional conditions. See e.g. Adams and Aizawa 1994, Loar 1991, and Seager 1993.

## REFERENCES

Adams, F. and Aizawa, K. (1994), 'Fodorian Semantics', in S. Stich and T. A. Warfield (eds.), *Mental Representation. A Reader.* Oxford: Basil Blackwell, pp. 223-242.

Block, N. (1986), 'Advertisement for a Semantics for Psychology', in P. French, T. Uehling and H. Wettstein (eds.), *Midwest Studies in Philosophy, vol. **10**: Studies in the Philosophy of Mind.* Minneapolis: University of Minnesota Press, pp. 615-678.

Brandom, R. (1994), *Making it Explicit. Reasoning, Representing, and Discursive Commitment.* Cambridge, Mass.: Harvard University Press.

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

Cummins, R. (1989), *Meaning and Mental Representation,* Cambridge, Mass.: MIT Press.

Cussins, A. (unpublished), *Constructions of Thought.*

Dretske, F. (1986), 'Misrepresentation', in R. J. Bogdan, ed., *Belief: Form, Content and Function,* Oxford: Oxford University Press. Reprinted in S. P. Stich and T. A. Warfield (eds.), *Mental Representation. A Reader.* Oxford: Blackwell, 1994. pp. 157-173.

Dretske, F. (1988), *Explaining Behavior. Reasons in a World of Causes.* Cambridge, Mass.: MIT Press.

Fodor, F. (1987), *Psychosemantics.* Cambridge, Mass.: MIT Press.

Fodor, J. (1989), 'Making Mind Matter More', *Philosophical Topics* 17, 1, pp. 59-79.

Fodor, J. (1990), *A Theory of Content and Other Essays*. Cambridge, Mass.: Bradford Books / MIT Press.

Haugeland, J. (1993), 'Mind Embodied and Embedded', in Y. Houng and J. Ho (eds.), *Mind and Cognition. Collected papers from 1993 International Symposium on Mind and Cognition*. Taipei, Taiwan: Institute of European and American Studies. Academia Sinica, pp. 3-37.

Israel, D. and Perry, J. (1990), 'What is Information?', in P. Hanson (ed.), *Information, Language and Cognition*. Vancouver B.C.: University of British Columbia Press.

Jackson, F. (1996), 'Mental Causation', *Mind* 105 (419), pp. 377-413.

Jackson, F. and Pettit, P. (1990), 'Causation in the Philosophy of Mind', *Philosophy and Phenomenological Research* 50, Supplement Fall, pp. 195-214.

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

Lepore, E. and Loewer, B. (1987), 'Mind Matters', *Journal of Philosophy* 84, (11), pp. 630-642.

Loar, B. (1991), 'Can We Explain Intentionality?', in B. Loewer and G. Rey (eds.), *Meaning and Mind. Fodor and His Critics*, Oxford: Basil Blackwell.

Loewer. B. (1994), 'From Information to Intentionality', in S. Stich and T. A, Warfuekd (eds.), *Mental Representation. A Reader*. Oxford: Basil Blackwell.

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

Millikan, R. (1984), *Language. Thought and Other Biological Categories. New Foundations for Realism.* Cambridge: MIT Press.

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

Neander, K. (1991), 'The Teleological Notion of "Function"', *Australasian Journal of Philosophy* 69 (4), pp. 454-468.

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

Seager. S. (1993), 'Fodor's Theory of Content: Problems and Objections, *Philosophy of Science* 60, pp. 262-277.

Segal, G. and Sober, E. (1991), 'The Causal Efficacy of Content', *Philosophical Studies* 63, pp. 1-30.

Smith, B. C. (1990), 'Putting Dretske to Work', in P. Hanson (ed.), *Information, Language and Cognition.* Vancouver B.C.: University of British Columbia Press.

van Gelder, T. (1993), 'The Distinction Between Mind and Cognition', in Y. Houng and J. Ho (eds.), *Mind and Cognition. Collected papers from 1993 International Symposium on Mind and Cognition.* Taipei, Taiwan: Institute of European and American Studies. Academia Sinica, pp. 57-82.