MEANING AND OTHER NON-BIOLOGICAL CATEGORIES

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

Abstract.

In this paper I display a general metaphysical assumption that characterizes basic naturalistic views and that is inherited, in a residual form, by their leading teleological rivals. The assumption is that intentional states require identifiable inner vehicles and that to explain intentional properties we must develop accounts that bind specific contents to specific vehicles. I show that this assumption is deeply rooted in representationalist and reductionist theories of content and I argue that it is deeply inappropriate.
I sketch the main features of plausible alternatives: such alternatives are either anti-representationalist (Dynamical Systems' models) or anti-reductionist (institution-based approaches), and are not committed to any such metaphysical premise.

1.   Introduction.

There are two important approaches in contemporary Philosophy of Language. One of them, let's call it the *formal* approach, is mainly concerned with those semantic properties of a given language that help us to understand the reference of its individual expressions and the truth-conditions of its sentences. This is semantics in its purest form. The other, let's call it the *psychological* approach, is likewise concerned with the semantic properties of the expressions of a language, but those semantic properties are now filtered through the psychological/biological features of the user of the language. Philosophers taking the psychological approach are interested in providing e.g. an account of the truth-conditions of the sentences of a language only insofar as such semantic analysis might help in the explanation of the linguistic and non linguistic behavior of the speaker. Semantic properties are thus the means, not the ends. The target is not so much to explain 'the relationships that hold between expressions of language and things in the world' as to explain why the expressions of a language have precisely the meaning they have and no other, and how the fact they have that meaning affects the behavior of the cognizer or language user. Alongside this psychological interest is often found an essentially *scientific* aspiration, namely, to reconcile semantic facts and a naturalistic world-view. Hence the proliferation of (among others) causal, informational and teleological accounts of content.

In this paper, I shall focus on the psychological, scientifically oriented, approach and, in particular, on teleological theories. A teleological theory of content is mainly concerned with the problem of content-fixing from a *naturalistic* point of view. The exact meaning of 'naturalism' in this context is itself an open question. The standard idea, however, is that semantic properties are not part of the primitive ontological furniture of the world and that, if we are to vindicate those properties as *real*, we must show how to unpack them into some other —more primitive and thus scientifically acceptable— set of properties. Naturalistic theories of content are

thus typically *reductionist* theories. The title of Ruth Millikan's influential monograph *Language, Thought, and Other Biological Categories* (1984) neatly encapsulates the spirit of such proposals. Language and thought are depicted as further parts of the natural order and meaning is reduced to a biological commodity.

I shall address the question of the *nature* of meaning against that teleological backdrop. My main contention is that the motivating spirit of teleological (and other naturalistic) approaches to meaning is deeply inappropriate precisely insofar as it seeks to characterize meaning in biological or evolutionary terms. In developing this thesis, I shall make explicit some metaphysical assumptions underlying teleological and, in general, naturalistic approaches to content. I will examine, in particular, the role played by the philosophical notion of *constitution* in the analysis of content and its consequences for a standard explanatory strategy in psychology. I will also develop some alternative metaphysical and explanatory notions which lead to a more positive thesis about the nature of meaning.

2.   Evolution and Constitution.

The semantic naturalist typically insists that an illuminating account of semantic properties be developed using the same type of explanatory framework found in the natural sciences, namely, a framework of causal relations and functional roles. Such a framework is supposed to account for the two most puzzling features of meaning: that meaning is *normative*, namely, that to know the meaning of an expression is to know how to use it correctly (even if that knowledge is not explicit), and that meaningful states, such as beliefs, can be false. Assuming that certain inner or outer tokens corresponding to a belief are produced if and only if a certain property is instantiated (e.g. there being bugs around), such belief tokens (or B-tokens for short) would be false if they are produced when that property is not instantiated. They would be used incorrectly in the same case. Add an evolutionary

spin to this causal approach and teleological accounts emerge. A given belief of type B has a particular content, e.g., 'there is a bug out there', if and only if some mechanism has been selected *because* (causally historically because) it has produced tokens of a certain type in the past whenever there were bugs around. This specific kind of naturalism with regard to intentionality thus involves a view of contentful mental states as states with biological purposes. Biological purposes are, in turn, characterized in evolutionary or selectionist terms.

I will take Millikan's approach as the paradigmatic teleological approach to semantics (But see also e.g. Papineau 1990 and Neander 1991). A detailed rehearsal of her work is beyond the scope of this paper, but what matters for our purposes may be summarized as follows: Mental representations are understood as 'teleofunctional items', i.e., as items which are produced by biological mechanisms that have been selected during evolutionary history and that are designed to perform some 'proper function'. Some item *A* has a function *F* as its proper function only if either:

> (1) *A* originated as a "reproduction" ... of some prior item or items that, *due* in part to possession of the properties reproduced, have actually performed *F* in the past, and *A* exists because (causally historically because) of this or these performances. (2) *A* originated as the product of some prior device that, given its circumstances, had performance of *F* as a proper function and that, under those circumstances, normally causes *F* to be performed by means of producing an item like *A*.
>
> (Millikan 1989a, p. 288)

The key to understanding representation is, however, to focus not on the representation's production, but on its *consumption*. The idea is that for something to be a representation at all, it must be a representation *for the system itself*. 'It is the devices that *use* representations that determine these to be representations and, at the same time ... determine their content' (Millikan 1989b, p. 282).

Only if certain conditions are met, can we say that a systemic mechanism will perform its proper function in consuming e.g. a belief token. Those conditions are (i) that a certain property P be instantiated when the system's mechanism consumes B-tokens and (ii) that the instantiation of that property actually explains why the production of B-tokens enables the system to perform its proper function. The content of a belief, like any other biological entity, thus turns not on the causes of the belief so much as on the advantageous results produced by the belief. Tokens of a certain belief type have as their content that e. g. there is a bug in front of the frog if it is required (for evolutionary purposes) that there be a bug in front of the frog in order for its bug detector neural mechanism to perform its proper function. In other words, we have to assume a

> certain rule of correspondence having the following property: unless we assume that some actual condition in the world corresponds in accordance with this rule to the representation confronted by the consumer, we cannot account, *with any single explanation that covers historical instances of consumer successes generally*, for *why* the consumer produces the effect that is its function.

> (Millikan 1990a, p. 154)

If, for the sake of the argument, you agree that e.g. frogs have beliefs, then frogs (who have a neural mechanism in virtue of which they snap their tongues when they detect a bug) have beliefs of the type 'there is a bug in front of me'. The content of the frog's belief is about bugs and not about e.g. visually identical black dots because the proper function of states of that type is related to the presence of bugs, not of black dots, i.e., because the neural mechanisms have been selected due to their producing tokens of that type whenever there were bugs around.

Millikan's position has been criticized in a variety of ways, of which I shall mention just two. In a very interesting paper

addressing a more general question concerning the consistency of the use of concepts with naturalistic views, Christopher Peacocke claims that teleological approaches are inadequate since they cannot offer a *constitutive* account 'of what it is to enjoy content-involving states' (Peacocke 1990, p. 63). Reformulating his contention using the frog example, Peacocke's diagnosis can be summarized as follows: teleological accounts are fine if what you want to explain is *why* the frog *has* a neural mechanism that produces tokens of the type 'there is a bug in front of me', but it doesn't help much with the *real* question, namely, *what it is* for the frog's neural mechanism to count as supporting a belief with that specific content.

Paul Pietroski has also criticized Millikan's teleological approach and has defended a thesis related to Peacocke's. The thesis is that the relation 'proper function' is *a* relation that beliefs can bear to properties sometimes instantiated in the system's environment. But 'proper function' is not, as Millikan claims, the relation 'has as its content that' (Pietroski 199, p. 268). The main reason for this mismatch is that, where intentional explanations are concerned, the emphasis should be on *discriminatory abilities* rather than on *historical importance.*

Pietroski diagnoses Millikan's failing as due to a conflation between *ethological* explanations (that show how a certain behavior plays a useful role in the life of a system or the members of a species) and *intentional* explanations (that show how a certain behavior is the product of how the system takes the world to be). Since both kinds of explanation sometimes make reference to the same properties, Millikan's teleological account of content appears plausible. But this overlap is just a fact that stands in need of explanation itself. What Millikan has done —according to Pietroski— is thus to sketch 'a new and potentially interesting kind of explanation of a certain range of behavior. But this is not intentional explanation; and the relation captured by Millikan's apparatus is not the relation that intentional states bear to their contents' (Pietroski 1992, p. 280).

Although Pietroski doesn't formulate his criticisms of Millikan's account in terms that involve the notion of *constitution*

(as Peacocke does), the conflation he mentions between Millikan's notion of proper function and the notion of content, and the conflation between ethological and intentional explanations could also be framed in such terms. What he is trying to show is that although Millikan's notion of proper function does capture properties that *contribute* to the individuation of the content of our beliefs, such properties don't *constitutively* characterize specific contents. That, at least, is how I interpret his main reason for claiming that explanations based on the notion of proper function don't always capture *what* it really *is* for the agent to have a belief with such-and-such a content.

Interestingly though, neither Peacocke nor Pietroski question the centrality of the notion of constitution itself. It is just taken for granted that constitution *is* the relation intentional states bear to their contents and that non-constitutive accounts of content ought to be dismissed as inadequate. I don't intend to discuss whether Peacocke's and Pietroski's analyses of Millikan's position are correct, but rather to examine their diagnoses of the problem. I find the diagnoses interesting in themselves because they raise two issues that deserve special attention. The first is whether the kind of relation intentional states bear to their contents must *necessarily* be a constitutive one. The second issue is what makes an intentional explanation a *good* explanation of some specific item of behavior.

3.   Metaphysical Commitments.

It is important to notice, first of all, that the whole teleological story only works on the assumption that mental states are *content carriers*: physical structures that encode information and that are *causally* correlated with instantiated properties. It is also important to realize that this same assumption lies at the core, not only of teleological approaches to content, but of a whole group of theories that can be gathered together under the rubric of representationalism.

Representationalism is, roughly, the positing of internal mental representations as the bearers of semantic properties, and it has been, at least until very recently (see below), something of an accepted dogma among philosophers of mind and cognitive scientists. Major differences concerned only the nature of the posited mental representations, with classicists and connectionists arguing in favor of language-like and distributed representations respectively. But this did not really affect the positing of internal representations themselves (even though the characterization of the appropriate notion of representation is not an easy philosophical task).

To remind you of just one aspect of that discussion, recall the occasional arguments supposed to establish the (non)representational nature of patterns of activation in connectionist systems. Distributed connectionist models are made up of networks of neuron-like units, and these units don't have a *direct* semantic interpretation. Since such networks' computations are completely determined by activity at the unit level, and the units each help to encode information of many different types, it is not possible to identify a stable and recurrent *entity* that corresponds to the classical notion of a symbol. There are thus no straightforwardly isolable vehicles at the physical level that can be identified as the articulated supervenience base on which the semantic properties of our mental states supervene. On this basis it has been argued that connectionist models of cognition cannot be used in support of the causal efficacy of content (Cf. Ramsey, Stich and Garon 1991).

The general reaction, however, was that this problem was more apparent than real. For a proper treatment of the connectionist paradigm —as a *cognitive model*— has to be developed at a level of description higher than that of mere numerical units and activations (see e.g. Smolensky 1988). It is not strange that, by restricting ourselves to the units and weights level, we get unsatisfactory accounts of *semantically* interpreted behavior. However, within the connectionist paradigm, there exist techniques of analysis that let us ascend to a higher level of description; a level

of description at which we *can* identify items that play the same role as the representations and rules of classical systems, although they have a character rather different to that of classic symbols and algorithms (Cf. Clark 1990).

Regardless of its own intrinsic interest, what I want to stress about this debate is, on the one (classicist) hand, the positing of semantic vehicles without further discussion and, on the other (connectionist) hand, the uncritical acceptance of the need to find the 'right level' of description: one where connectionist analogues to classic semantic vehicles might be identified. Both positions thus maintain a deep commitment to the idea of content carriers whose physical properties can enter in causal explanations of behavior.

This Aristotelian image of the content and the container imports with it an implicit thesis about the kind of relation that exists between intentional states and their contents. The thesis, roughly put, is that such a relation is to be cashed out in terms of a notion of *constitution*. The idea is that what it is for a particular thing to be a thing of a certain kind is completely determined by the stuff of which it is made, by its intrinsic properties. Intrinsic properties make something what it is and are thus *constitutive* of that particular thing. Applying this general metaphysical idea to the intentional realm results into the following deep framing assumption: what makes an intentional state have the content it has are the properties constitutive of the physical state upon which such intentional properties supervene. This notion of constitution stands out as the metaphysical glue that binds intentional states to their contents.

To focus the overwhelming influence of this seldom discussed metaphysical assumption (an exception is Rudder Baker 1995), consider a different and important debate in contemporary philosophy of language and philosophy of mind: the debate between internalism and externalism. Internalism is the view that holds both that the content of a mental state supervenes on intrinsic physical states of the subject, and that such contents are individuated 'narrowly', i.e., without essential reference to the subject's physical and social environment (see e.g. Fodor 1975, 1980). The externalist

position, by contrast, denies that mental contents supervene on intrinsic physical properties of the subject. The externalist claims that contents are individuated 'widely', i.e., by reference to the subject's environmental or social context (see e.g. Burge 1979, 1986). The point I want to stress is that the main challenge for the externalist is to show that relational properties can actually play a *constitutive* role in the characterization of the contents that putatively explain the agent's behavior. The internalist can agree that the individuation of computational states of cognitive systems is relative to the properties of the world that they inhabit. But, from such an internalist point of view, this is *just* to say that the world *determines* the agent's cognitive states (via the inputs), and not that the external features can play any *constitutive* role in the characterization of such states. The externalist has to show how her claims about content can be *constitutive* claims, and fans of externalism usually try to do so without ever questioning the plausibility of the basic requirement.

Why should this notion of constitution be lauded as the only acceptable metaphysical relation for the characterization of content? The root of the answer lies in a widely accepted conception of the mental: a conception that involves the combination of representationalism and a computational view of the mind. Once we commit ourselves to this computational representationalism, we commit ourselves to the idea of inner vehicles of mental content. Semantic properties are *contained* in some physical vehicle that acts as the guarantee of their causal efficacy, and thus allows the philosopher to reconcile the semantic realm and the natural world. But, if we opt for such a (reductionist) strategy, then what drives the characterization of content is the internal complexity of the state, i.e., its intrinsic, constitutive properties. Additional pressure comes from the idea that computational explanations fix the form of *good* psychological explanations. Since computations, at least as standardly conceived, cannot be sensitive to properties external to the system, we seem forced once again to the conclusion that the only properties that count in binding the intentional realm and the

natural world are the intrinsic properties of constituting physical states.

It is therefore an interesting philosophical exercise to look at a non-representationalist, non-computationalist approach to cognition to see how things work there. If we find that non-representationalist, non-computationalist accounts can actually provide good intentional explanations, then we will be able to conclude, at least, that that familiar notion of *constitution* doesn't *necessarily* have to be the only acceptable metaphysical relation that might bind intentional states to their contents, and therefore that broadly non-constitutive accounts should not be rejected solely due to their non-constitutive status[1]. The Dynamical Systems approach to cognition, is one of the few exceptions in contemporary Cognitive Science to both the dogmas of representationalism and of computationalism. Although a detailed discussion of the Dynamical Systems' viewpoint is beyond the scope of this paper, I shall sketch some of its main features in the next section. In Section 5, and with a more philosophical eye, I will sketch some pragmatic accounts of content which place the ontological notion of *institution* (as opposed to constitution) at the core of their analysis.

4.   The Dynamics of Cognition.

Cognitive Science is largely agreed on at least this: that at the heart of a scientific understanding of cognition lies the construct of internal representation. Both defenders of explicit, syntactically structured, classic representations (e.g. Newell & Simon 1972; Fodor & Pylyshyn 1988) and of connectionist, distributed representations (e.g. Smolensky 1988) retained the fundamental idea of inner computational states acting as the vehicles of specific contents. Cognitive Science is also agreed on the idea that computational explanations are good intentional explanations and that a computational theory of cognition (one which builds in the notion of representation. The two theses are thus internally related) is the correct way to go about the explanation of cognitive behavior.

In recent years, however, a much more radical view has gained some ground. This view calls into question the commitment to internal representation itself and the suitability of computational theories of cognition as explanations of the behavior of cognitive agents (see e.g. Beer 1995). Cognition, it is claimed, need not involve the creation and manipulation of anything deserving the name of 'internal representations' at all. The image of 'cognition as computation' is said to be no longer the 'only game in town' (van Gelder 1995, p. 346). Under this new perspective, cognition is characterized as 'state-space evolution in certain kinds of non-computational dynamical system' (van Gelder 1995, p. 345).

Without going into too much detail (see Clark and Toribio 1994 for an extended discussion), the general idea is this: internal representations are an unnecessary and unproductive posit for the analysis of real-time, environmentally-embedded behavior. The dynamicist thus stresses the difference between systems which are intrinsically computational and systems which merely admit of computational modeling (e.g. planetary motion). The difference is between systems which really do have internal representations and system which do not (Cf. Beer 1995). Cognitive systems, it is claimed, are better *characterized* in terms of continuous circular exchanges between the physical / biological features of an organism and those of the environment in which the organism is embedded and functioning. Where content-involving descriptions are suitable for the explanation of such a cognizer's behavior, they involve a notion of content that is characterized by the abilities of the systems to interact in specific ways with the world in which they are situated.

The relevant systemic properties are thus not constitutive properties in the classical, standard sense, i.e., they are not intrinsic, context independent properties of the system whose behavior we trying to explain. Instead, the properties that count for the characterization of content are, mainly, external, context- and environment-dependent properties of the system. This is because the kind of behavior that mainly interests the anti-representationalist is characterized by constant circular causal

exchanges between the system and the environment in which the system is embedded.

As a result, the notion of internal content *carriers* has vanished from the (pure) Dynamical Systems approach. Instead what we find is a general formalism based on the idea of an abstract geometric state space whose dimensionality is of arbitrary size (depending on the number of relevant system parameters). The system's behavior is explained in terms of location and motion within that space. Some geometrical constructs are invoked, constructs that capture properties of certain points or regions in the space. The mathematics typically specifies a dynamical law which determines how the values of a set of state variables evolve through time. Given an initial state, the temporal sequence of states determined by the dynamical law constitutes one trajectory through the space. The set of all the trajectories passing through each point is called the flow, and the shape of this 'flow' is the typical object of study (see e.g. Abraham and Show 1992 for more details).

Nothing in this picture preserves the idea of a particular computational state as the vehicle of the content that explains a given behavior of the system. And without the idea of a 'vehicle', we lose also the idea of the intrinsic properties that constitute it. Yet the Dynamical Systems literature is full of cases in which we obtain nonetheless a good explanation of a target system's behavior. This is especially clear in developmental applications. Here, Dynamics is emerging as the most suitable explanatory framework in which to address e.g. processes of self-organization and adaptation underlying the development of mature cognition. Increasingly psychologists are seeking to understand the changes involved in cognitive development in dynamical instead of representationalist terms. This non-computational approach has been most extensively pursued, it is true, in cases where the parameters to which the system is responding are straightforwardly physical and are available to the systems as proximal stimuli. This is the case, for example, in explanations of reaching and grasping in infants (see e.g. Thelen 1995). But the model has also begun to be applied to highly complex tasks such as natural language processing. Jeffrey

Elman, for instance, has developed a model of natural language processing in which the ability to predict the next word in a sentence after having being presented with the first *n* words in a sequence is analyzed using dynamical tools. Here, the idea of locations in a state space replaces the idea of internal representations of words, the structure of that space replaces the idea of a dictionary or word meanings and the dynamics of the system replaces the idea of grammatical rules (see Elman 1995).

Of course, it could be argued that even a given flow in an abstract geometrical space is *constituted* by such-and-such points or trajectories, and that such constitutive properties of the flow are the relevant ones in the explanation of the system's behavior. Such a response, however, misses the main point of the alternative approach, because the set of trajectories passing through each point wouldn't be as it is without the continuous reciprocal interactions that obtain between the system itself and various external features. Such constant reciprocal causal interaction is built into the dynamical description of the state space. Features external to the biological organism thus figure in the account of cognition itself. In this vein, van Gelder and Port comment that:

> Cognitive processes span the brain, the body, and the environment; to understand cognition is to understand the interplay of all three. Inner reasoning processes are no more essentially cognitive than the skillful execution of coordinated movement or the nature of the environment in which cognition takes place. The interaction between "inner" processes and "outer" world is not peripheral to cognition, it is the very stuff of which cognition is made.
>
> (Port & van Gelder 1995, p. ix)

The answer to the 'real' question about content, namely, what it is for the system to have a belief with such-and-such content is thus answered, within this dynamical view, in terms that invoke the system's *abilities to act* in the larger environment in which is situated[2]. Those abilities are not what they are merely in virtue of

some underlying intrinsic physical structure. Instead, what matters are rather the 'superstructural' facts concerning the space of environmental interactions into which the system enters. The metaphysical relation that applies here is thus not that of simple constitution. The relevant relation is one of *institution* (more on which below).

An important caveat. My depiction of the Dynamical Systems approach as anti-representationalist should be understood against the background of a view of representationalism as roughly equivalent to the belief in inner computational vehicles for specific contents. I do not thereby preclude some alternative thesis about the nature of representation itself. Obviously, how we define representation will greatly affect our views about the full force of the 'anti-representationalist' claims (see Clark & Toribio 1994). But given a familiar and restricted interpretation of representationalism, the idea is clear: we are to reject inner computational states as the key explanatory factor in understanding intentional behavior, and with it the metaphysical assumptions concerning constitution. Constitution by intrinsic physical or functional states should not be assumed as the clear criterion for any naturalistic account of content.

5.　Cognition Institutionalized.

Let me now make some related points at a more philosophical level. Some recent theories about content do take, as their core explanatory notion, the practical interactions of the subject with her environment, especially those kind of interactions that involve some discriminative ability on the part of the subject. Such discriminative abilities, it has been suggested, are critical to the project of individuating the contents of our thoughts. The claim is not new. It can be found in some of the writings of the later Wittgenstein and is central to Dummett's and Evans' characterizations of the nature of a theory of meaning (See Dummett 1975, 1976; Evans 1982). More recently, Adrian Cussins' (1992, ms.) and Robert Brandom's (1994)

analyses of linguistic content stand out as existence proofs of the potential richness of this alternative explanatory strategy. Both these projects defy easy rehearsal, but I'll attempt a brief sketch of what is most important for our purposes.

Brandom's project is to develop a pragmatic account of the content of the expressions of a language in terms of a notion of commitment. Commitments are characterized by a kind of deontic status and deontic statuses are *instituted* (not constituted) by social practices. The ontological notion of institution 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 ... ... The explanatory strategy adopted here is to begin with practices that institute deontic statuses of commitment and entitlement and then to show how those practices thereby confer specifically propositional conceptual contents on what is assertible.
>
> (Brandom 1994, pp. 48, 169)

Cussins' project is to develop what he calls a 'naturalized transcendentalism', But what is most interesting for our purposes is the fact that institution —even if not as explicitly as in Brandom's case— again lies at the core of the posited relations between intentional properties and physical properties. Material, social and historical contingencies institute the complex network of normative activity systems that make human cognizers what they are:

> 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 ms., p. 4)

In Section 2, we saw how teleological views —and naturalistic views, in general— try to account for the two most interesting facts about meaning: its normativity and the fact that meaningful states can be false. We saw that such accounts include a reductionist thesis (in the case of teleology, the reduction is toward biologically characterized functions), and a causal explanatory strategy based on constitutive physical properties. The key move in both Brandom's and Cussins' proposals is rather to make normativity the central, irreducible feature of meaning. The goal of accounting for semantic properties in a naturalistic way thus becomes the goal of preserving that essential feature. As a result, the standard naturalistic strategy is altered in many respects. First, no reductionist assumptions underlie this way of characterizing semantic properties. The normativity of linguistic content is the starting point and is treated as ineliminable. Second, although the normative dimension of language is claimed to be ineliminable, it can still be *explained* and an explanation of it is offered, so it is not treated as some kind of bare primitive. Third, this explanation is not framed in causal or constitutive terms. The normativity of language is understood as *instituted* by social-practical activity:

> Though this normative dimension of linguistic practice is taken to be ineliminable, it is not treated as primitive or inexplicable. It is rendered less mysterious in two ways. First, linguistic norms are understood as *instituted* by social-practical activity ... The second way norms are rendered less mysterious is 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, pp. xiii-xiv)</div>

Once again, the 'real' question about content, the question about what it is that confers a particular content on an intentional state, is answered in terms of the superstructures into which the biological system enters. An additional advantage of the more philosophical treatment is that, unlike the Dynamical Systems accounts, it is especially designed to encompass more abstract properties of the environment (the aspects of the social setting can be as rarified and distal as one desires), and it is especially geared towards an account of linguistic meaning. In fact, social institutions are the super-structures *par excellence* for the characterization of linguistic meaning. To know how to use a linguistic expression correctly requires membership in a social institution and the normativity of meaning (the basic feature that more naturalistic approaches were unsuccessfully trying to capture) is intrinsically and essentially attached to such institutions.

A word of warning. I do not claim to have shown that accounts built on the idea of institution offer better theories of content (although I try to defend that thesis elsewhere. See Toribio submitted). What the alternative institutional view brings to light here is just the possibility of characterizing the notion of content in a deeply non-reductionist manner. It is thus an existence proof of the possibility of an institutionalized account of intentionality: an account in which meaning may be analyzed in many different ways but in which it is *not* a biological category.

If we now recall our earlier discussion of teleology, the situation looks rather interesting. For it now turns out that physical constitution is not necessarily the metaphysical glue that binds intentional states and their contents. It follows that to say (as e. g. Peacocke does) that teleology does not offer a *constitutive* account of content is not yet to provide a good reason to reject Millikan's account as inappropriate. If we still believe such an account to be incorrect, we must develop a different argument. This is the task of the next section.

6. Intentional explanations, teleological explanations and good explanations.

The general commitment to physical-constitutional explanations is usually tied up with the idea that good intentional explanations ought to be *causal* explanations. This link is not accidental. Indeed, intentional explanations are characterized in causal terms only on the assumption that what causes something to have a particular effect is nothing but its physical-constitutive properties. As we saw in Section 2, the internalist position in the internalism / externalism debate is a good example of this connection in action. The justification of the internalist's view flows from the physicalist bias of standard scientific methodology, namely, the claim that the causal powers of any event are completely determined by its local physical features

In teleology, however, causation is not all. There is, in addition, a historical, selectionist ingredient. Causation thus becomes *historical* causation and good intentional explanations are a species of *teleological* explanation. Intentional states have the content they do 'causally historically because' the mechanisms that produce the content carrying states have been selected due precisely to their possession of those semantic properties. The addition of this historical component to the causal explanatory framework inherited from the naturalistic / reductionist trend creates, I shall now suggest, an important tension within the teleological picture. On the one hand, attributions of beliefs to a system are said to be true in virtue of how the system is historically related to its physical environment. This permits a characterization of content in teleofunctional terms and seems to allow the teleologist to distance herself from any strong version of computational representationalism. After all, it is the history of an item that determines its proper function rather than the item's present properties (Cf. Millikan 1989a). It would thus seem misleading to construe teleology as involving the physical-constitutive thesis, i.e., as committed to the idea that what counts for explaining intentional properties are just local physical features,

viz. the inner vehicles of content. Such a construal would go against the official thrust of the teleological approach, which is to insist:

> that psychology is not at root a science involving laws, that explanations in psychology are unlike explanations in the physical sciences, that it inescapably is a deeply ecological science dealing with how the organism interacts with its wider environment.
>
> (Millikan 1993, pp. 11-12)

Yet, residual (and I think distortive) features of the 'inner vehicle' model remain in place. The teleologist still takes contentful mental states to provide reasons that *cause* the agent to behave one way rather than another. And theorists such as Millikan are explicitly committed to the idea that beliefs have map-like inner vehicles. On her account, a system's inner mechanisms can perform their proper functions only because such inner maps are in place, even if the mapping rule (the rule that, under normal conditions, determines what is represented) is fixed by the evolutionary history of the system (Cf. Millikan 1986, p. 78).

We are thus left with a rather uncomfortable situation. A teleologist like Millikan maintains much of the apparatus springing from the metaphysical commitments of more standard naturalized approaches, including the idea that there must be real inner vehicles of content; vehicles that play a causal role in the explanation of the behavior of the system. But, at the same time, what determines that those inner vehicles have the particular contents they do is not their present properties but the evolutionary history of the mechanisms that produce them.

This residual tension would be removed if the teleologist were to contend that what matters for the characterization of content is the interaction of the *whole* organism with its environment rather than any properties —historical or otherwise— of the organism's *internal mechanisms*. But to opt for such a solution is effectively to give up the idea of e.g. inner maps whose causal powers explain the

system's behavior. In other words, such a solution would involve the outright rejection of the guiding scientific ideas underlying computational representationalism. Yet such ideas look to be practically ineliminable from Millikan's account, which still displays a clear representationalist/reductive character. To reject such ideas would be to turn the usual teleological approach to semantics into a much more radical theory. Nonetheless such a change wouldn't necessarily be *bad news*. In fact, it looks to be the easiest way to relieve the tension and to absolve Millikan's account from the criticisms noted in Section 2.

Even such a liberalized version of the teleological approach, however, faces deep (and I believe fatal) difficulties. The deep trouble is that such theories simply fail to provide a good theoretical basis for the scientific explanation of intentional behavior. The reason is that 'evolutionary accounts ... will be unable to distinguish between causal-explanatorily equivalent content attributions' (Godfrey-Smith 1994, p. 273). A small fable, due to Paul Pietroski, makes the point well:

> The kimus live near a large rocky hill. Their only predators are snorfs, carnivores who roam past the hill each morning. Kimus used to be "color blind". But in virtue of a genetic mutation, one particular kimu —call him Jack— came to have an internal mechanism M that produced tokens of a physically specifiable state type B in the presence of certain wavelengths of light. Each morning, something red on the hilltop caused Jack to form a B-token when he looked up. And Jack (like his descendants) turned out to have a "fondness" for red things; i.e., other things being equal, Jack would move towards the distal causes of B-tokens when such tokens were produced. So each morning, Jack trudged up the hill and thereby avoided the snorfs. Natural selection took over; and Jack's mechanism type proliferated throughout the species. There was no other reason (e.g., detection of food) for the selection in favor of having the "color mechanism".

> (Pietroski 1992, p. 273)

Millikan claims that, for a kimu, a B-token 'signifies roughly, "fewer snorfs this way"' (Millikan 1990b, p. 149). According to Pietroski, however, '[t]here are several candidates for the content of a (current) kimu's B-token: Lo, redness; Lo, Wavelength W; Something nice is over there; There's that nice mountain top again; etc. Or perhaps there is no determinate content of B-tokens to speak of. But ... B-tokens are not about snorfs' (Pietroski 1992, p. 273). Given that kimus don't have any means of recognizing snorfs at all, and given that they can discriminate red things from non-red things but can't reliably discriminate snorfs from non-snorfs, it is surely implausible to claim, as Millikan does, that 'kimus go up the hill because they believe that there are no snorfs there' is a good intentional explanation.

A further reason for doubting the teleological account derives from the character of psychological explanations themselves. If we opt for a teleological characterization of content, we commit ourselves to the idea that the environmental properties that count for the explanation of an agent's behavior are always 'real' properties of the world. Like Millikan, we must invoke a 'rule of correspondence' that reflects some 'actual condition in the world' and a condition that affects the individual, not only as an individual, but as a member of one particular species. Such commitments, however, go against the usual explanatory strategy in psychology. In general, psychological explanations —explanations that invoke intentional states of one kind or another— do not appeal to evolutionary considerations in order to account for our behavior. The explanatorily central issues in psychology do not turn on debates concerning the evolutionary proper function of mental states. One reason is that an agent's intentional states do not have to be tied to the objective conditions for her survival. In fact, they don't have to be tied to objective conditions at all! Much of our mental life involves tracking properties of objects in such a way that the internal representations of those objects are not representations of the objects *themselves*, i.e., of the objects taken in any objective way at all. Instead, it is only the *mode of presentation* of the object

under a particular perspective and within a particular behavioral context that counts. By trying to characterize the contentful intentional states that appear in psychological explanations in teleological terms, we end up focusing on properties that might be very important from a biological point of view, but that don't seem to be that relevant in the explanatory context of psychology. Psychology is interested in explaining behavior (in explaining e.g. why an agent has responded in such-and-such way to a particular situation). But it is not clear that the ability to token a particular thought or the ability to respond appropriately to a particular situation must afford an agent some evolutionary advantage over other members of its species. And even if there were such evolutionary advantage, it is not obvious how to include it as a relevant explanatory parameter in the psychological context.

To sum up, I have tried to display a general metaphysical assumption that characterizes basic naturalistic views and that is inherited, in a residual form, by their leading teleological rivals. The assumption is that intentional states require identifiable inner vehicles and that to explain intentional properties we must develop constitutive accounts (history-based or otherwise) that bind specific contents to specific vehicles. This assumption, I have argued, is deeply rooted in representationalist and reductionist theories of mind and content. However, there exist plausible alternatives that due either to their anti-representationalism (Dynamical Systems' models) or their anti-reductionism (institution-based approaches) are not committed to any such metaphysical assumption. I have sketched the main features of these alternative views, stressing the fact that they seem nonetheless to provide good explanations of intentional facts. I claim that the existence of such counterexamples casts real doubts on the need to invoke constitution as the necessary metaphysical glue binding intentional states and their contents. A side-effect, however, is that some standard criticisms of teleological approaches —criticisms that accuse them of not offering a full or constitutive account of content— no longer apply. The real problem with such teleological views, I finally argue, is not to be found at the metaphysical so much as at the explanatory level. Teleological

approaches fail because evolutionary considerations simply don't reflect the explanatory interests that prevail in psychology. If meaning is a biological category, then that category doesn't belong to the domain where intentional explanations matter, i.e. to psychological explanations. We all know how to apply a *modus tollens*.

References

Abraham, R. H. and Shaw, C. D. 1992: *Dynamics. The Geometry of Behavior*, 2nd ed. (Redwood City: California: Addison-Wesley).

Beer, R. 1990: *Intelligence and Adaptive Behavior* (Academic Press, San Diego).

Beer, R. 1995: 'A Dynamical Systems Perspective on Environment Agent Interactions', *Artificial Intelligence* 72, pp. 173-215.

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

Burge, T. 1979: 'Individualism and the Mental, *Midwest Studies in Philosophy, Volume IV* (Minneapolis, MN., University of Minnesota Press), pp. 73-121.

Burge, T. 1986: 'Individualism and Psychology, *The Philosophical Review* 95, pp. 3-45.

Clark, A. 1990: 'Connectionist Minds', *Proceedings of the Aristotelian Society* 90, pp. 83-102.

Clark, A. and Toribio, J. 1994: 'Doing Without Representing?', *Synthese* 101, pp. 401-431.

Cussins, A. 1992: 'Content, Embodiment and Objectivity: The Theory of Cognitive Trails', *Mind* 101 (404), pp. 651-688.

Cussins, A. ms.: *Constructions of Thought.*

Elman, J. L. 1995: 'Language as a Dynamical System' in R. F. Port and T. van Gelder (eds.) *Mind as Motion. Explorations in the Dynamics of Cognition* (Cambridge, Mass.: MIT Press), pp. 195-225.

Fodor, J. 1975: *The Language of Thought* (New York: Thomas Y. Crowell).

Fodor, J. 1980: 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology'. Reprinted in J. Fodor, 1981: *Representations* (Cambridge, Mass.: MIT Press), pp. 225-253.

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

Fodor, J. & Pylyshyn, Z. 1988: 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition* 28, pp. 3-71.

Godfrey-Smith, P. 1994: 'A Continuum of Semantic Optimism', in S. Stich & T. Warfield (eds.) *Mental Representation. A Reader* (Oxford: Basil Blackwell), pp. 259-277.

Millikan, R. G. 1984: *Language, Thought and Other Biological Categories. Foundations for Realism* (Cambridge, Mass.: M.I.T. Press).

Millikan, R. G. 1986: 'Thoughts without Laws', *The Philosophical Review* 95 (1), pp. 47-80. Reprinted in R. Millikan, 1993.

Millikan, R. G. 1989a: 'In Defense of Proper Functions', *Philosophy of Science* 56, pp. 288-302. Reprinted in R. Millikan, 1993.

Millikan, R. G. 1989b: 'Biosemantics', *Journal of Philosophy* 86 (6), pp. 281-297. Reprinted in R. Millikan,1993,.

Millikan, R. G. 1990a: 'Compare and Contrast Dretske, Fodor and Millikan on Teleosemantics', *Philosophical Topics* 18, pp. 151-161. Reprinted in R. Millikan,1993.

Millikan, R. G. 1990b: 'Clarifications on *Language, Thought, and Other Biological Categories*', *Annals of Scholarship* 7, pp. 147-149.

Millikan, R. G. 1993: *White Queen Psychology and Other Essays for Alice* (Cambridge, Mass.: MIT Press).

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

Newell, A. & Simon, H. 1972: *Human Problem Solving* (Englewood Cliffs, NJ: Prentice-Hall).

Papineau, D. 1990: 'Truth and Teleology', in D. Knowles (ed.) *Explanation and its Limits* (Cambridge, Cambridge University Press), pp. 21-43.

Peacocke, C. 1990: 'Content and Norms in a Natural World', in E. Villanueva (ed.) *Information, Semantics, and Epistemology* (Oxford: Blackwell), pp. 57-76.

Peacocke, C. 1994: 'Content, Computation and Externalism', *Mind and Language* 9 (3), pp. 303-335.

Pietroski, P. M. 1992: 'Intentionality and Teleological Error', *Pacific Philosophical Quarterly* 73, pp. 267-282.

Port, R. F. and van Gelder,T. (eds.) 1995: *Mind as Motion. Explorations in the Dynamics of Cognition* (Cambridge, Mass.: MIT Press).

Ramsey, W., Stich, S. and Garon, J. 1991: 'Connectionism, Eliminativism and the Future of Folk Psychology' in Ramsey, W., Stich, S. and Rumelhart, D. (eds.) *Philosophy and Connectionist Theory* (London: Lawrence Erlbaum), pp. 199-228.

Rudder Baker, L. 1995: *Explaining Attitudes. A Practical Approach to the Mind* (Cambridge: Cambridge University Press).

Smolensky, P. 1988: 'On the Proper Treatment of Connectionism', *Behavioral and Brain Sciences* 11, pp. 1-74.

Thelen, E. 1995: 'Time-Scale Dynamics and the Development of an Embodied Cognition', in R. F. Port and T. van Gelder (eds.) *Mind as Motion. Explorations in the Dynamics of Cognition* (Cambridge, Mass.: MIT Press), pp. 69-100.

Toribio, J. submitted: 'Naturalizing Normativity', *Australasian Journal of Philosophy.*

van Gelder, T. J. 1995: 'What Might Cognition Be if Not Computation?', *Journal of Philosophy* 92 (7), pp. 345-381.

Wilson, R. A. 1994: 'Wide Computationalism', *Mind* 103 (411), pp. 351-372.

Footnotes

1   An alternative is to argue for a different notion of computational explanation, one which still involves a constitutive view of the relation that intentional states bear to their contents, despite being characterized in externalist terms (see e.g. Peacocke 1994 and Wilson 1994). Such an alternative, however, is not so interesting for my purposes here, as it doesn't question the centrality of the notion of constitution itself.

2   As we saw in Section 2, Pietroski's criticisms of Millikan were rooted in the relevance of present discriminatory abilities to good intentional explanations. Unfortunately he didn't seem to realize that such vindication involved not only a rejection of history, but also a rejection of some deeper metaphysical theses.