Mental Representations and Millikan's Theory of Intentional Content: Does Biology Chase Causality?*

Robert D. Rupert

In her landmark book, *Language, Thought, and Other Biological Categories* (Millikan1984),[1] Ruth Garrett Millikan utilizes the idea of a biological function to solve philosophical problems associated with the phenomena of language, thought, and meaning. Language and thought are activities of biological organisms, according to Millikan, and we should treat them as such when trying to answer related philosophical questions. Of special interest is Millikan's treatment of intentionality. Here Millikan employs the notion of a biological function to explain what it is for one thing in nature, a bee dance (43), for example, to be *about* another, in this case, the location of a nectar source. My concern in this paper is to understand whether Millikan's account of intentionality adequately explains how humans achieve reference, in language or thought, to individuals and groups in their environment.

In bringing her theory of intentional content to bear on human activities, Millikan focuses largely on natural language. Thus, in what follows, I begin by laying out the biology-based principles that underlie Millikan's theory of content, then proceed with an explanation of how the theory is to apply to natural language. As it appears, Millikan's account of how content is determined for natural language terms and sentences rests on the determinacy of intentional content at the psychological level. This leads me to take a careful look at what Millikan says about the content of mental representations, in hopes of finding a sufficient basis there for the application of Millikan's theory of content to natural language.

Ultimately, I conclude that Millikan's theory faces a problem of vacuity. If we approach the theory as a theory of intentional content, intended to explain the nature of reference, the theory is lacking in an extremely important respect: Millikan explains how it could be one of the biological functions of a mental or natural language term to refer, without telling us precisely what in the natural order constitutes the reference relation. It is one thing to say that *x*'s

biological function is to refer; it is another to explain what reference is. My conclusion will take some getting to, however, and I plan to argue by way of identifying an apparent circularity in Millikan's theory of content. This circularity is important in its own right, but it is also illuminating in that it would seem best rectified by supplementing Millikan's theory with a theory of content of an entirely different sort, one that is causal, covariational, or informational in nature. By identifying the circularity, then, we should be able to see more clearly what Millikan's theory of intentional content lacks.

Before proceeding to the details of Millikan's theory, some preliminaries are in order regarding the theoretical framework in which the discussion of mental terms is set. First, without intending to make too many commitments regarding the nature of the mental medium, I frequently refer to mental terms as terms in a language of thought (or LOT, for short).[2] Further, I give *terms* in LOT as much attention as I do for two general reasons, which have no special relationship to Millikan's theory: I sometimes focus on terms for the simple reason that in the development of theories of content for LOT, many philosophers have done so.[3] Also, I am inclined to think that the content of a thought is partly a function of that thought's components, where we might reasonably think of some of these components as terms: mental content seems subject to certain principles of compositionality.[4]

While these two reasons for focusing on the reference of terms may have some force in their own right, this paper is supposed to be about Millikan's theory of content. And given that Millikan is not among the ardent defenders of a language-like medium of thought, nor, for that matter, of principles of compositionality, the approach that I've proposed may seem a bit wrong-headed. For example, Millikan says, "I will argue that mental sentences, *should they occur*, would be intentional icons. But it seems quite reasonable that much of our active thinking, much of our inferring, may occur in media more like maps or models (in the lay sense) than like sentences." (Millikan 1993a, 114, emphasis added) In the same piece Millikan expresses reservations about the idea of mental processes as computations over mental sentences (Millikan 1993a, 104). Comments such as these seem to distance Millikan from hard-line LOT

enthusiasts.

On the other hand, Millikan makes it quite clear that mental representations, even maps or models, must contain the equivalent of referring terms (71, 96, 139-40, 241-4, 1993a, 113-4, 1993b, §10 and 11): for beliefs to play their role in reasoning, beliefs representing premises must share terms that have the same referents; this is so whether we're talking about beliefs explicitly represented in a quasi-linguistic system or sensory-based beliefs embodied in a map-like medium. In the former case, note that to reason through a valid categorical syllogism, you must rely on a middle term. This term appears once in each belief that corresponds to a premise and only plays its proper role in reasoning if it is treated as having the same reference in both its appearances. In the case of sensory-based beliefs, shared reference effects the coordination of data coming from different senses, which coordination is necessary to put the sensory data to practical use: normally when a subject ties a knot, she succeeds in doing so by treating certain elements of her tactile image as having the same referent as corresponding elements of her visual image (240). Millikan's commitment to individual mental terms is also made clear when she claims that at least some thoughts must have subject-predicate structure (and be subject to negation; Millikan 1993a, 117-118, 1993b, 106).

Millikan is cautious: she wants to leave open the possibility that mental representations do not possess the degree of structure possessed by natural language sentences, while yet claiming that the *contents* of mental representations imply the natural language sentences that accurately characterize such contents (Millikan 1993a, 119-120). But regardless of the precise form our mental representations take, if they are to guide inference and action in the way Millikan claims, they must contain referring elements; we can call these elements 'terms in LOT', for want, perhaps, of a superior locution.

I. Millikan's Theoretical Framework

According to Millikan, the intentional content of a sign is determined by the proper function of the sign. Defined generally, the proper function of *x* is the function of *x* the successful performance of which is responsible for the continued reproduction of members of the reproductively established family of which *x* is a member (28). Millikan defines proper function in this general way to allow other things besides symbols (e.g., hearts) to have proper functions. The precise definition of 'reproductively established family' is fairly technical (see 27-28, for a full definition), but the idea is one which can be easily conveyed by examples. A reproductively established family can be, among other things, a species, a bodily organ type, a syntactic structure type, or a mental state type.[5]

An explanation of how a proper function has been performed in a way that leads to continued reproduction of members of a reproductively established family is called a 'Normal explanation' of the performance of the proper function (33-34). It is important to recognize that 'Normal' is not meant in the statistical sense: a Normal explanation is not simply an explanation of how the mechanism in question *usually* performs; a Normal explanation explains how the mechanism performs in cases where the performance actually contributes to reproductive success. (Millikan also employs the term 'Normal conditions' to refer to the conditions which must hold in order for a proper function to be successfully carried out in accordance with a Normal explanation.) Sometimes a device or organism can have a proper function that it successfully performs only rarely, statistically speaking, yet the successful performance of which is key to the continued reproduction of the device or organism. Millikan gives the example of sperm. The chance that any individual sperm will ever fertilize an ovum is very, very small. And historically speaking, of all sperm which have existed, the percentage that have fertilized ova is tiny, yet it is still the proper function of any individual sperm to fertilize an ovum. It is only because sperm have fertilized ova in a sufficient, albeit comparatively small, number of cases, that sperm, as a type, has continued to exist.

The distance between statistical normalcy and biological Normalcy also appears in the psychological realm. So, for example, even though the majority of desires may go unfulfilled

(though it is hard to be sure of the numbers here), the proper function of a desire is "to produce a state of affairs onto which it maps in accordance with certain mapping rules" (140). As a reproductive family of psychological states, desires have continued to be reproduced in humans because in a sufficient, though perhaps relatively small, number of cases, they have successfully brought about of the state of affairs they describe (or map onto).[6]

As we will see in the following sections, the way in which the proper function of a symbol, or 'sign', to use Millikan's terminology, determines the intentional content of that sign can be fairly complicated. Central to Millikan's theory is the claim that the proper function of a sign is determined by the way in which the sign assists an interpreting device in the performance of its (the interpreting device's) own proper function. The role of interpreting devices, described as such, seems to be clearest when we think of natural language signs. For this reason, among others, I now turn to Millikan's application of her theory of intentional content to natural language.

II. Sentence Content and Term Extension

A. Term Extension in a Natural Language

According to Millikan, sentence content is primary and the content of terms secondary in natural languages:

> The most basic or most direct kind of correspondence, then, is the correspondence between a true sentence and a world affair. When this correspondence occurs, we say that the sentence has a "real value" -namely, the affair it maps onto. A less direct, more mediated, kind of correspondence is the correspondence between a referential term *in the context of a true sentence* and its referent. And riding piggyback upon this second kind of correspondence is an even more derivative kind of correspondence--the correspondence of a lone term to its referent or of a term in a

> false sentence to its referent. Indeed, this last kind of correspondence is of a *totally* different kind...It is, roughly, the relation that one thing has to another qua being, only, *supposed* to correspond to it. (104)

The extension of a term (i.e., what the term corresponds to or is supposed to correspond to)[7] is secondary to sentence meaning, but it is not independent of sentence meaning. According to Millikan, term extension is determined by the real values of the set of true, token sentences in which the term has appeared. In what follows, I examine in more detail the two most important aspects of this process by which sentence meaning determines term extension. First, given that the extension of a term is determined by the set of *true* sentences in which the term appears, we need to know which sentences are true, according to Millikan: we need something like a semantics for natural language sentences. Second, we want a precise description of the relation between sentence content and term content: we must know in what specific way the extension of a term is derived from the content of the true sentences in which the term has appeared.

A sentence is true, according to Millikan, when it corresponds to, or maps onto, the world (103-104). Whether or not a sentence S maps onto the world is determined by S's mapping rules (107-108), which are themselves determined by a complex pattern of historical relations. This pattern of historical relations should be seen as function taking the four following arguments: (1) all of the historical uses of S, (2) all of the historical uses of transforms of S (where a transform of a sentence is its negation or is the result of making a substitution in the sentence--acceptable substitutions being ones that replace a term from a given syntactic category with a term from the same syntactic category), (3) the Normal conditions for the performance of the proper functions of S, and (4) the Normal conditions for the performance of the proper function of transforms of S.

Millikan's idea is that when a sentence is changed into one of its transforms, the state of affairs that is the Normal condition for the performance of the proper function of the sentence (the ‘real value’ of the sentence) changes also. "It is *transformations* of the icon that correspond

to *transformations* of the real value." (107) These transformations determine variants and invariants in respect to aspects of the sentences and aspects of the states of affairs that are the Normal conditions for the performance of the proper functions of the sentences. Take, for example, "John flies" as a transform of "John swims". The invariant term in the set consisting of these two sentences is the term 'John'. In the states of affairs that are the Normal conditions for the performance of the proper functions of these two sentences, the invariant aspect is John. Therefore, 'John' maps onto John.

Now we can see more clearly the relation between sentence content and term extension in Millikan's theory. A term's extension is given by the mapping rules of the sentence in which the term occurs. The mapping rules for the sentence spell out (or 'articulate') the correspondences between the parts of the sentence and the parts of the state of affairs onto which the sentence maps, for example, the correspondence between 'John' and John, when 'John' appears in a true sentence. It's important to keep in mind that the mapping rules for a given sentence are not built up out of mapping rules for the individual constituent terms. The Normal conditions for the fulfillment of *sentences'* proper functions are the initial data driving the determination of sentences' mapping rules. As individual elements, variants and invariants emerge from the Normal conditions for the performance of the proper functions of a variety of sentences. This I cannot stress enough: the nature of these Normal conditions is an historical matter; questions about them can only be answered by discovering the Normal conditions for the performance of the proper functions of the sentences that contain the terms in whose extensions we are interested. This leads to a question of paramount importance, "What are the Normal conditions for the performance of the proper functions of sentences?" In particular, what are the Normal conditions for the performance of the proper function of indicative sentences, their being the ones that can have truth-values?

The proper function of an indicative sentence in a natural language, qua sentence of the indicative mood, is to bring the *hearer* into a belief state that accurately reflects a relevant state of affairs in the world. If there had not been a sufficient number of indicative sentences with this

effect, hearers would have stopped paying attention to indicative sentences, and speakers would have stopped producing them (58-59, 31, 99). Thus, the creation of a true belief in the cooperating interpreter, i.e., the hearer, is the ultimate test whether an indicative sentence has performed its proper function.

It seems, then, that for Millikan, assignment of content to any sentence or term in a natural language depends (metaphysically, not epistemologically) on the determinacy of belief content; there must be at least enough determinacy for some of these beliefs to qualify as true beliefs. For a natural language sentence S to have content, there must have been historical situations in which S fulfilled its proper function (or at the very least, there must be other sentences, S1....Sn, that contain the constituent terms of S, which themselves, i.e., S1...Sn, performed their proper functions). This much is clear. But since the proper function of any indicative sentence is to produce a true belief in the hearer (58), S and its transforms can have fulfilled their proper function only if past hearers were in a position to have true beliefs. For this reason, my exposition of Millikan's theory of content for natural language sentences would seem to be lacking an essential component. Without some determinate conditions for a belief's being true, there is nothing to separate successful past uses of S, the ones that contribute to the determination of S's mapping rules, from the unsuccessful uses. Thus, it is (nearly) time to turn to Millikan's treatment of beliefs.

B. Alternative Interpretations Considered

Founding a theory of natural language sentence content on a theory of true belief causes a serious problem for Millikan, to be described in the sections below. Given the seriousness of the problem, it seems worthwhile to consider the possibility that I have misrepresented Millikan's views. At certain points, Millikan seems to imply that a natural language sentence S's mapping rules are determined independently of the truth-conditions for beliefs, even that the content of a belief B is dependent upon the content of the natural language sentence that caused B. In the

first half of what follows, I consider two passages that suggest this alternative interpretation of Millikan; and in each case I argue that the passage in question does not justify ascribing to Millikan the view that S's mapping rules are determined independently of the existence of belief content. Though the correct interpretation of individual passages is sometimes less than obvious, I proceed to argue that there is a decisive, general reason to reject interpretations according to which S's content is determined independently of belief content. I close this section by describing and rejecting a final attempt to avoid my favored interpretation of Millikan's theory of content. According to this final alternative, sentence content and belief content are interdependent in the sense that both sentence and belief content emerge together from an analysis of human communication and behavior as a whole.

Millikan says, "If sentences are intentional icons, this is not *because* they translate into beliefs and intentions which are themselves intentional icons—although, I will claim, they do so translate. It is because they display the four characteristics of intentional icons I have laid down." (101, emphasis original) From the sound of this passage, belief content is not the basis of sentence content. However, the sound becomes fainter when one reads the passage keeping clearly in mind Millikan's assiduous and, I believe, successful attempt to distance her theory from a Gricean theory of natural language meaning (Chapter 3). According to a Gricean theory, the content of a sentence in a natural language is a function of the contents of speakers' intentions when they use such sentences: if English speakers usually use 'Snow is white' with the intention of conveying the idea that snow is white, then 'Snow is white' means that snow is white. In contrast, Millikan's four content-determining conditions do not found meaning directly on speakers' intentions. Millikan's point seems to be that regardless of what other factors may come into play, intentional content is ultimately determined by the four conditions she has laid down. These conditions make the determination of content depend on the performance of the relevant proper function(s) of the interpreter of an intentional icon (97-99). According to Millikan's theory, then, if we are curious about the relation between sentence content and belief

content, we must begin by identifying the relevant proper functions of sentences and beliefs. But as I argued in the preceding section, once Millikan has identified the proper function of indicative sentences, the proper function of a given indicative sentence can be fulfilled only if some beliefs in past hearers possessed determinate truth-conditions independently of natural language sentence content.

A second, potentially problematical passage seems to suggest that I have reversed the direction of the dependence relation between belief content and sentence content. In discussing the relationship between public language and mental sentences, Millikan says, "Any such inner sentence is supposed to end up being used to adapt my thinking processes or my behavior to whatever conditions the outer sentence mapped." (151) Here it sounds as if the outer sentence already maps onto the world before it causes the subject to token an inner sentence, the inner tokens inheriting their content from the public language sentence that caused the tokening of the inner sentence; however, it is essential that we place this passage in context: we must be sure to understand what "such inner sentence[s]" are for Millikan.

At this spot in her exposition, Millikan has taken up the question whether in learning language, all competent speakers come to associate with a given natural language term a definite, shared intension (150). For Millikan the inner sentences under discussion here are simply internalized copies of natural language sentences, phonetically isomorphic to the public language originals. And in talking about what such an inner sentence is 'supposed to' do in the passage quoted above, Millikan describes the proper function of the mechanism that generates internal copies of public language sentences during the language learning process--before, as we might put it informally, the hearer even understands what the sentence means: the function of this mechanism is to create internal structures that can eventually be used as a means to respond appropriately to natural language sentences.

It would be a mistake, then, to think that Millikan has here made a sweeping claim about the relationship between thought and language; for learning a natural language, and learning to associate some mental sentence with a natural language sentence, is *not* the same thing as

learning to think or coming to have beliefs. Though you may sometimes think in what seems to be an internal copy of your natural language, thinking is not the same thing as using an inner voice to talk to yourself (At the very least, there is no reason to attribute to Millikan the view that this is the nature of human thought). It should be clear from our discussion in the introductory section that Millikan is hesitant to attribute the precise structure of a natural language to human thought; and in her discussions of human activities guided by sensory input, she endorses the existence of mental representations that are not explicitly coded as linguistic analogues of natural language sentences (240). That there are such non-linguistic representations illustrates the limited scope of her discussion of the intensions associated with natural language sentences. Thus, when Millikan seems to suggest that inner sentences inherit mapping rules from the mapping rules of the natural language sentences that cause the inner sentences, we should see this as a specific point about learning to associate an intension, shared by other speakers, with a given natural language sentence or term, not as a point about the very nature of LOT or what it is to have a belief. Millikan's claims about how we learn a shared natural language should not be taken to bear on how the contents of beliefs, generally speaking, are determined.

It may bolster my argument to note the nature of the intensions of which Millikan speaks. These are not extension-determining intensions of the sort often discussed in philosophy of language; for Millikan, mapping rules determine extension, and she refers to the mapping rules as the 'senses', not the intensions, of linguistic elements (111). In contrast to sense, an intension is, according to Millikan, an internal routine governing the iteration of the linguistic element associated with that intension (133); and as will be discussed in more detail later, such inner term iterating mechanisms serve their proper function when they help the subject to acquire a set of accurate beliefs. Millikan's point in the quoted passage can thus be explained in the following way: We know that an indicative sentence has as its proper function to cause a true belief in the hearer; as a result of doing so in the past it has acquired its mapping rules. Bearing this in mind, consider an internal sentence S* that has been copied from indicative natural language sentence

S as part of the language learning process; S* has as its function to ensure that S causes in the budding language user the same true belief whose being caused by S in the past accounts for S's mapping rules. This is the sense in which the inner sentence inherits its content from the natural language sentence's existing mapping rules. We must not forget, though, whence those mapping rules originally came, i.e., the causing of true beliefs in past hearers.

While some readers may remain unconvinced by the preceding discussion, a consideration of the larger theoretical picture unequivocally supports my preferred interpretation of Millikan. The alternative interpretations are unsatisfactory because they leave a significant gap in Millikan's theory, a gap that she seems to have no intention of leaving unfilled: on the alternative interpretations, there is no explanation of whence come the mapping rules for a given natural language sentence S. In contrast, I have taken at face value Millikan's description of how the mapping rules for a given indicative natural language sentence S are determined. Mapping rules for S are determined by certain past uses of S and certain past uses of transforms of S. Not every past use of S (or of other sentences containing linguistic elements of S--henceforth I omit this qualification when talking about past uses of S) bears on the question whether a correspondence exists between elements of S and elements of the world. We only take into consideration past uses of S in cases where these uses facilitated the performance of the proper function of the appropriate interpreting device. If I say "Snow is white", and the hearer wiggles and giggles and proceeds to shoot himself (too much philosophy of language, I would guess), we might well suspect that the interpreting device in the hearer has not performed in accordance with its proper function. Thus, this situation would seem to have no bearing on the determination of mapping rules for the English sentence 'Snow is white'. Neither are mapping rules determined by a straightforward mathematical function of all past uses of S and the situations of those uses. I can sit in my room all day speaking out a series of English sentences, systematically transformed, while at the same time making systematic but odd changes in my environment, without this in any way affecting the mapping rules for the sentences I utter along the way (At least, let's hope that Millikan's theory does not imply that the mapping rules would

be so affected). It is key to Millikan's theory that the mapping rules for S are determined by a restricted class of past uses of S and its transforms (This is one reason why it is so important to distinguish between Normal conditions and statistically typical conditions).

What, then, is the origin of the mapping rules for S? When laying down the determinants of mapping rules, Millikan states as her first condition that "the real value [the situation onto which the intentional icon maps] is a Normal condition for performance of the icon's direct proper functions." (107) But what are direct proper functions of the icon? If the icon is an indicative natural language sentence, then its proper function is to create true beliefs in the hearer. "[T]he stabilizing function of the indicative mood is to convey information, i.e., to produce *true* belief." (53-54, emphasis original) (The stabilizing function of the indicative mood is the function "performed in a critical mass of cases of actual use" and which accounts for the continued use of the mood [31-32]). An easy interpretation of Millikan's theory is that the mapping rules for S are determined just by whatever situation is such that without its presence, S could not do its communicative job. However, if this interpretation is to be taken seriously, we must be given some way to identify the situation necessary for S to serve its communicative function. Millikan plainly describes what is required of an indicative sentence in order for it to continue to be produced and used. "Clearly, the hearer listens [to an indicative sentence], then uses a certain rule of interpretation in forming a belief only because these acts have correlated in the past with the formation of *true* beliefs." (58, emphasis original) We should not saddle Millikan's theory with incompleteness when she herself recognizes the need for filling in and offers the requisite details.

In the face of the preceding, some may hold out hope for an interpretation of Millikan according to which the mapping rules for S are determined independently of the creation true beliefs. Consider yet another passage suggestive of this view. Millikan says:

> "In the case of the indicative intentional icons, the Normal explanation of how the icon adapts the interpreter device such that it can perform its proper functions

> makes reference to the fact that the icon maps onto something else in accordance with a specific mapping function of a kind to be described below" (99)

It sounds as if there is a fact as to what the icon maps onto independently of what the interpreter device does or doesn't do. Applied to the case at hand, one might read this passage as saying that the mapping rules for S exist prior to the formation of true beliefs in any hearer and that the mapping's existence is a condition for an interpreter's coming to have a true belief on the basis of hearing S; yet reading the passage in this way would again create for Millikan the problem of incompleteness that she does not seem to think her theory faces: she would be left without an explanation of whence come these independent mapping rules for S. In the quoted passage, Millikan leaves open the specific nature of the mapping function. But once the nature and the determinants of the mapping function have been spelled out for indicative natural language sentences, we find a theory of content of the sort I've been describing. According to this theory, we are free to think of a *current* use of S as causing a belief whose truth conditions match the existing mapping rules for S; however, we must not lose sight of the source of those mapping rules. As I have argued based on ample textual evidence, S's mapping rules emerge from those past uses of S that have fulfilled their proper function of creating true beliefs in hearers, in which case the determinacy of the current mapping rules for S depends on S's having caused hearers to have true beliefs in the past. This in the *only* source of natural language mapping rules for indicative sentences that Millikan describes explicitly.

Millikan's theory of natural language sentence content appears to rest on a separate semantics for belief. As a final attempt to avoid this commitment, consider the following, alternate explication of Millikan's theory, an explication that may show how it can be the function of the indicative form to produce true beliefs without this requiring independent specification of what it is for a belief to be true. According to this approach, we should focus on the *cooperative* relationship between symbol producer and symbol interpreter. We are to see these components as working in concert, as a complete system or as parts of some larger system,

and we ask what function each component plays in the overall system. On this view, the assignment of functions to parts of the overall system does not proceed in a building-block fashion: we do not begin by assigning a function to one component of the system, and then assuming determinacy of the first assignment, use that assignment as the basis of a function assignment to some related component, and so on. Instead, we uncover functions, of the entire system as well as its components, by empirical theory construction. We observe the system as a whole and assign functions to its parts in an attempt to formulate the best explanation of how the entire system functions successfully (e.g., by surviving). No single component's function *defines* another's. When we apply this view to the topic of linguistic content, we see the producer and the consumer of an intentional icon as working together to facilitate the performance of each other's proper function, qua linguistic producer or consumer.

According to the presently suggested view, my demand for an explanation of what it is to have a true belief is illegitimate. I have made it sound as if Millikan simply cannot proceed with her discussion of sentence content until she has provided us with conditions for the determination of thought content; but, the objection runs, my mistake is in thinking that we must completely define the function of the interpretive device (the belief forming mechanisms) before we can give an account of what it is for the sentence producing device (the speaker) to succeed at its task.[8]

So far as I can tell, there is nothing wrong with assigning a general function to sentences by decomposing the overall system of language users and consumers into components and assigning functions to the components on the basis of the way they contribute to the overall success of the system (or to the success of various other components of the system). Furthermore, I do not take anything I have said so far to be at odds with such an approach. Complications arise, though, when we move beyond the general attribution of a function to indicative sentences and attempt to understand how the content of specific sentences can be determined in accordance with that function. Millikan claims that the proper function of the indicative mood is to produce true beliefs in the hearer. To understand this claim, it is not necessary that we have in mind a

detailed semantic theory for LOT. However, to apply Millikan's theory to any specific, indicative, natural language sentence S, we must ask whether specific beliefs were caused by specific past occasions of S's use, and we will have to know which of those specific beliefs were true; we cannot do so without having in hand some fairly specific semantic theory for beliefs. Further, regardless of what semantic theories we happen to have developed, there cannot be a fact of the matter as to which past beliefs were true *unless beliefs have their truth conditions determined independently of the truth conditions for S*; but again, this is not to say that we cannot understand in some general way the overall functions of beliefs and sentences, as types of things, by functionally decomposing the system(s) in which such items are found and by which they are used.

Note that my demand for an independent determination of truth-conditions for beliefs rests in no way on the assumption that Millikan is offering us analytic *definitions* of terms such as 'indicative mood'. I do not suppose Millikan to be giving us the results of her conceptual analysis of the idea of the indicative mood; nor do I suppose her to be offering a completely isolated explanation of the proper function of the hearer's indicative-sentence processing mechanisms. Millikan simply offers what seems to be a very sensible suggestion as to the function of the indicative mood, i.e., to convey information to hearers, or in other words, to cause the hearers to have true beliefs. Millikan arrived at this view, I would guess, on the basis of systemic considerations, including considerations of the behavior of humans, their acts of communication, their natural environment, and their social organizations.

Perhaps a pair of examples here will help to illustrate my point. Imagine that you stand in front of a table and say, "There's a table", and the result is that I form a true belief. On Millikan's view, a third party who assumes that everything has gone Normally should be able to go back and figure out what you were referring to with your use of 'table'; it must have been the table. The assignment of the table as the referent of 'table' best explains why I developed a true belief in response to your saying "There's a table".[9] However, this process can only proceed as Millikan describes it if the third party knows *whether or not I developed a true belief* in response

to your saying ‘table’, and more importantly (to leave the epistemological mode), the third party only gets the *right* answer *if there is a determinate matter of fact as to whether I formed a true belief* in response to the utterance of “There’s a table”.

Contrast human natural language with Millikan’s oft-used example of the bee dance. In the overall scheme of bee life, the bee dance has the function of leading bees to nectar. Naturalists discovered this through careful empirical observation. This function identified, naturalists then attempted to figure out what relation would most likely have to have existed between the elements of bee dances and elements in the environment in order for bee dances to have succeeded in getting bees to nectar sources in those cases where the dances did so. How does the bee dance compare to the case of indicative sentences? First off, note that in the case of bee dances, the consumers’ proper function, qua consumers of bee dances, has nothing to do with bee *beliefs*. The proper function of the relevant interpreting device is to get the watching bees to the nectar. Millikan repeatedly uses the bee dance example to illustrate her theory of proper functions; in all cases, her emphasis is on the movement of the dance-consuming bees toward the source of nectar. For example, “it is a proper function of the interpreting devices in the watching bee to produce a direction of flight that bears a certain relation to this dance.” (40-1; see the preceding and following pages for similar descriptions; also see 1993b, 99-100). And further:

> Notice that the device that interprets an intentional icon does *not* always do this by producing *another* intentional icon. For example, although the mechanisms that interpret bee dances “translate” these dances into a direction of flight of the bee, they do this only in the physicist’s sense of “translate”. The result of this “translation” is a bee flying in a certain direction. But a bee-flying-in-a-certain-direction is not an intentional icon, for there is no cooperating device that interprets it. (101)

And in what comes immediately after, Millikan discusses the contrasting case of sentences,

noting, among other things, that the performance the proper function of sentences does involve the creation of intentional icons, i.e., beliefs, in the hearer, Thus, when comparing the case of bee dances to that of indicative sentences, we should bear in mind the proper functions of the relevant interpreting devices: in the former case, to get the bees to nectar, and in the latter case, to cause the hearer to have true beliefs.

At the most general level, the function of the bee dance and of human linguistic communication is to help the relevant organisms survive. However, such a function is far too indeterminate an affair for its successful performance to fix the intentional content of a specific bee dance or indicative sentence. This is, presumably, one reason why Millikan identifies, in both cases, a more specific function as the source of intentional content. If bee dances seem to have easily identifiable, determinate content, it is because the proper functioning of the interpreting device seems easy to check on. While such seemings may be deceiving, it seems to be a cut and dried matter: either the bees fly toward a nectar source upon watching the dance, or they do not.[10] Millikan owes us a theory of belief content largely because matters are not cut and dried in the case of true beliefs. We may have a general sense of what a true belief is, i.e., a belief that corresponds to reality (but even this is controversial); however, in identifying the presence of a true belief in a human, there seems to be nothing analogous to the bee's success in flying toward or reaching a nectar source, at least not from the epistemological standpoint.[11] Metaphysically speaking, it may be a perfectly determinate matter whether a given indicative sentence causes a hearer to have a true belief. But if we are after an illuminating account of the semantics of indicative sentences of natural language, we will be unsatisfied with an account that gestures toward the intuition-based attribution of true beliefs. If causing true beliefs is the proper function of the indicative sentence, and this proper function is the basis for the content of the sentence, then we should only be satisfied by a detailed explanation of the source of a belief's truth conditions, in virtue of which it is true or false in a given situation.

I propose, then, to take seriously Millikan's claim that the mapping rules for S are determined by S's history of facilitating the performance of the proper function of the

interpreting devices, i.e., S's (and its transforms') history of creating true beliefs in hearers. This view comes directly out of the text, and so far as I can tell, Millikan offers no other view as to how to isolate those past utterances of S and its transforms that count toward the determination of S's mapping rules. Therefore, it is to a consideration of belief content we now turn.

C. Belief Content and LOT Term Extension

In developing a semantics for natural language, Millikan emphasizes the relation between speakers and hearers: speakers produce sentences that fulfill their proper functions by aiding hearers in the fulfillment of the hearers' proper functions. When considering the meaning of mental terms, however, one is tempted to think that the dynamics of the situation *must* have changed. No one hears me think, and thus the semantics of LOT should be determined independently of any cooperating interpreter. To the contrary: Millikan's theory of LOT content structurally mirrors her theory of intentional content for natural language sentences, the difference being that the sign producer and the cooperating interpreter reside in the same person. The producers are numerous; they are what Millikan calls 'inner-term iterating programs' (144). Such programs consist of testing routines that determine whether or not a subject will mentally token a given term of LOT: the routines range from brute sensory processes to consciously applied criteria for the application of a term.

What about an interpreter? In general, the mapping value of an indicative intentional item is determined by the role the intentional item plays in those cases where the cooperating interpreter fulfills its proper function qua interpreter of intentional items of the type in question (99). At the receiving end of indicative sentences of LOT, Millikan claims that the relevant interpreter device is a consistency tester, a device that tests the outputs of the term-iteration programs against each

other for consistency. Thus, in order to figure out what determines the extension of terms in LOT, we must first develop a clear understanding of the proper function of the consistency testing device.

The consistency tester is a functionally cohesive, but not necessarily spatially localized, device whose proper function is to test the consistency of the various sentences produced by LOT sentence and term production mechanisms. In Millikan's words, "S's relevant interpreter [where S is a sentence in LOT] is the consistency tester qua tester of *other* programs that can produce tokens of the same type as S or negations of these." (146) The proper function of the consistency tester is to create a consistent set of representations of the world, or more accurately, to produce a set of sentence and term producing programs that generate a consistent set of representations of the world in LOT.

Note that the proper function of the consistency tester cannot be to test sentences for consistency in the strictly formal sense. Millikan says that in order for the consistency tester to perform its proper function, sentences in LOT must map onto the world in at least some cases (146). But there is no reason why any LOT sentence must map onto the world in order for the consistency tester to test a subject's set of beliefs for *formal* consistency. Say that a subject mentally tokens 'S' and 'It's not the case that S'. Imagine, then, that the subject's consistency tester sets off the inconsistency alarm. If the consistency tester's proper function is only to test for formal consistency, then the consistency tester has done its job; the catch is that it has done so regardless of what connections might exist between S and the world, and regardless of whether such connections were the same for both tokenings of S (one being S and the other tokening occurring in "It's not the case that S").

The merely formal test for inconsistency can not be what Millikan has in mind as the proper function of the consistency tester. Instead the relevant function of the consistency tester would seem to be that of testing sets of beliefs for truth. Presumably, the consistency tester continues to be reproduced because, given the proper background assumptions, testing for consistency helps subjects maximize the amount of truth in their belief sets (13, 143, 324, and Chapters 16

and 17, passim). A consistency tester would seem to be of little evolutionary value to someone in a world in which testing for consistency did nothing to help increase the amount of truth in her set of beliefs; but it is useful in the world assumed by Millikan, a world in which knowledge of the environment keeps you alive and in which consistency is a mark of truth in your beliefs about that world.

Utilizing Millikan's theoretical terminology, we should say that testing for truth in a belief set is the *focused proper function* of the consistency tester. A focused proper function is the proper function of a device that lies at the end of a series of functions and which is such that "the device doesn't (properly) do much else, except as this else *depends* upon the completion of the last of these stages or of the stage where these functions converge." (37) Thus, testing for formal consistency may be a proper function of the consistency tester (because if it were not for the ability of the consistency tester to test for formal consistency in a critical number of cases the consistency tester would not be reproduced), but testing for formal consistency is not the focused proper function of the consistency tester. Testing for formal consistency is only a stage in a series of functions that culminates in the test for truth, after which the truth serves a plethora of purposes. The test for truth is the focused proper function of the consistency tester, because it is only after truth has been successfully located that the output of the consistency tester can fulfill the numerous functions (e.g., helping us to find food, shelter, and mates) that diverge thereafter and all depend on the identification of truth.

Millikan illustrates her concept of a focused proper function with the example of the human eye (37). This example provides a nice analogy to the case of the focused proper function of the consistency tester. The eye does all sorts of things that we place under the rubric ‘the mechanics of vision’. However, these mechanical processes are useful only because they lead to the production of visual representations. After the point at which visual representations are produced, these visual representations are used "for anything one happens to need them for." (37) The consistency tester is similar to the eye. The test for formal consistency is one of the processing steps that leads up to the ultimate output of the consistency tester (together with

adjunct devices), the production of a set of accurate representations. Once we have a set of representations, taken to be true, we use these representations in all sorts of ways.

Millikan claims that the LOT term production programs must produce terms that map onto the world in a consistent fashion in order for the consistency tester to perform its proper function as a cooperating interpreter (146). If S is 'My local grocery store only sells red apples', then the specific variants in S (e.g., 'apple') should map onto the world in the same way as those variants do when they appear in other LOT sentences. This makes perfect sense when we understand the function of the consistency tester as a test for truth rather than mere formal consistency. If Millikan were to deny that the consistency tester has as its proper function the production of *accurate* representations, while still maintaining that the consistency tester is the appropriate interpreter for the various LOT term production programs, it would be unclear why it mattered whether 'apple' referred to apples on one occasion of its use and skyscrapers on another. In contrast, the process of testing for formal consistency does increase the likelihood of the subject's acquiring true beliefs assuming 'apple' consistently maps onto apples. The question remains, however, "What determines the mapping rules for LOT sentences, from which the assignment of reference to individual terms is to follow?"

Here a deep problem arises for Millikan's theory of intentional content, a problem in the theory's application to LOT *as well as to natural language sentences*. Begin with natural language. The mapping rules for a natural language sentence S are determined by the way in which S (and its transforms) have facilitated the formation of true beliefs in past hearers. This leads us to consider the content of beliefs. As for any intentional icon, the content of a belief should be determined by the way it facilitates the performance of the proper function of its cooperating interpreter. However, given that the consistency tester is the relevant cooperating interpreter for beliefs, and given that the focused proper function of the consistency tester is to provide the subject with true beliefs, there is no non-circular way to apply Millikan's theory of content to LOT sentences: Millikan is committed to saying both that (1) the consistency tester's past successes in performing its proper function establish the mapping rules for indicative LOT

sentences (i.e., beliefs) and (2) the consistency tester's past successes were the cases where its test for consistency created true beliefs. But how can the consistency tester succeed in creating true beliefs unless the truth conditions for the relevant beliefs have been previously and independently established? Here we find the circularity promised at the outset. The consistency tester's past successes determine the mapping rules (thus the intentional content) of LOT sentences; but for the consistency tester to have had a past success it must have created a true belief; however, to have created a true belief requires the mapping rules for that belief to have already been in place. (What else might Millikan have in mind as a condition for a belief to be true, other than for it to map accurately onto the world?) There can be no true beliefs in the absence of determinate mapping rules, yet an application of Millikan's theory of content to LOT seems to tell us that in order for mapping rules to be determined, the subject must have had some true beliefs in the past. Thus, on pains of circularity, the consistency tester's past successes in helping the cognitive system to construct true beliefs cannot be a determinant of the mapping rules for beliefs.[12] Millikan's theory of intentional content seems to face a grave difficulty, not one that can be easily dealt with by making minor adjustments in the theory: Millikan offers no alternative suggestion as to how mapping rules for beliefs might be determined, and it's not clear how, given her various commitments, Millikan might provide such an alternative.[13]

A theory of content is unacceptably circular if it claims that the semantic values of the relevant terms and sentences are determined entirely by the previous appearance of true sentences of the same type. Millikan's theory tells us that the mapping rules for LOT sentences are a function of the past successes of the consistency tester. However, there can only be such past successes if the mapping rules for LOT sentences are defined; the consistency tester can enjoy no success without determinate conditions for true belief being in place. For this reason I find Millikan's theory of intentional content for LOT terms and sentences unsatisfying.

We can encapsulate Millikan's theory of content in the following principle:

P1: The content of sentence S = the state of affairs in the world that was

> essential to a Normal explanation of past cases when a standardized interpreting device performed its proper function in interpreting S.[14]

I have argued that P1 is circular when applied to LOT. We should find the circularity especially troubling when we recall that for P1 to apply to natural language in the way Millikan intends it to apply, the semantics of LOT must itself be determinate. Hearers of natural language sentences must be capable of having true beliefs in order for them to perform their proper function qua interpreting devices of indicative, natural language sentences; but in order for the hearers to have true beliefs, there must first exist a semantic relationship between their beliefs and the world. Therefore, if the circularity identified above inhibits the application of P1 to LOT, it also undermines Millikan's theory of natural language content.

III. Intentional Content and Representational Content

How might Millikan's theory avoid this apparent circularity? In the presentation of her theory, Millikan distinguishes between intentionality and full-blown representation (71, 96). Thus far, I have focused only on Millikan's theory of intentionality. Given that we normally think of sentences and terms as representations, it would seem compulsory that I examine closely Millikan's theory of representation.

According to Millikan, a representation is an intentional item the referents of the elements of which are supposed to be identified by a cooperating interpreter. Thus representations are a species of intentional items, a species that includes sentences in natural language and LOT. Pointing this out, however, does nothing to alter the preceding discussion. Consider the sort of identification Millikan has in mind in the case of LOT sentences (touched on above, in the introductory section): when two LOT terms that have the same extensions appear in two different LOT sentences from which an inference is to be drawn, the relevant, internal interpreting device can only perform its proper function if it treats these two different terms as

though they have the same reference (241-2 and 1993b). For example, whether or not the term takes precisely the same form in both of its appearances, the middle term of a categorical syllogism must be treated as having the same reference in the major and minor premises in order for the inference to make sense (It would seem arbitrary otherwise). The same holds true when we combine data from two different sensory systems to coordinate action or when we combine current the data from one sensory system with memories (240-1). The consistency tester, too, engages in acts of identification. In order to do its job, it must recognize various tokens of the same LOT term type as such, *and* it must treat instances of two *different* LOT terms as having the same reference when they do in fact map onto the same variant in the world (Millikan gives precedence to the latter case in 1993b).

While provocative, Millikan's explanation of what it is to be a full-blooded representation is of no help in resolving the problem I have raised for her theory of intentional content. Saying that two terms are treated as having the same reference does nothing to tell us what determines their actual reference (or intentional content, or real value); Millikan's remarks concerning what is distinctive of representation simply do not speak to the determination of actual reference. Though we can understand why it is useful for the consistency tester, for example, to treat certain pairs of terms as co-referring, we understand this only because we assume that the two terms do, in fact, refer to the same thing; and it is the theory of intentional content that is supposed to explain how such reference is fixed in the first place. Thus, Millikan's theory of representation is of no help in removing the apparent circularity from her theory of intentional content.

IV. The Consistency Tester and Causal Relations

The apparent circularity of Millikan's theory results from the failure to describe the proper function of the consistency tester in non-semantic terms. But while she says what seems to be too little about this issue, we may be able to locate in Millikan's remarks a plausible, non-

semantic characterization of the consistency-tester's proper function. When discussing the intentionality of beliefs, Millikan characterizes the consistency tester as a tester of programs for the production of sentences in LOT:

> [These programs] should pass muster only if they are helping to produce sentences that map onto the world in accordance with some definite rules *for a reason*--a reason mentioning conditions under which the programs often operate and mentioning laws of nature which, under these conditions, *connect* these sentences with what they map. (146)

This passage suggests that the proper function of the consistency tester is to make sure that the LOT sentence-producing programs stand in the correct *causal* relation to the world. On this reading of Millikan, the proper function of the consistency tester *is* described non-semantically. This non-semantic characterization secures the possibility that the mapping rules for sentences in LOT are determined in a way analogous to the way Millikan claims that the mapping rules are determined for natural language sentences: Take a sentence S of LOT. The mapping rules for S are determined by the history of S's tokening. In particular, S maps onto the state of affairs that was present in past cases where the tokening of S was used by the consistency tester to successfully perform its proper function of testing the relevant LOT sentence-producing programs for the proper causal relation of S (or its elements) to the world.

Now Millikan's task is to explain the nature of the causal relationship for the presence of which the consistency tester is supposed to provide a test. Of the sentences produced by the sentence producing programs being tested, Millikan says that "these sentences must bear *information* concerning what they map onto roughly in the sense that Dretske defines in *Knowledge and the Flow of Information*." (146; see Dretske 1981) Thus, the causal relation for which the consistency tester tests might best be seen as the relation 'bears-information-about'. This would solve the problem of circularity by specifying a causal relation the testing for which

is the consistency tester's focused proper function: the consistency tester's focused proper function is no longer the semantically-laden function of increasing truth in the belief set; instead it is to maximize the amount of information, characterized non-intentionally, carried by LOT term- and sentence- tokenings. It is by the acquisition of information that the deliverances of the consistency tester can be used successfully for their various purposes.

Assume that the focused proper function of the consistency tester is to test for the causal relation 'bears-information-about'. We should worry, then, that Millikan has not given us a theory of intentional content for LOT terms or sentences beyond her deference to Dretske. If the truth conditions of beliefs are ultimately determined by the existence of the causal relation of bearing information, then we must focus on this relation if we are to understand intentional content as a natural phenomenon.

Millikan has recently commented on the relation of her theory to other prominent theories of content, including theories of an informational or indicational nature:[15]

> Picturing, indicating, and inference are equally involved in human representing, but as biological norms rather than as mere dispositions. It is not the facts about how the system *does* operate that make it a representing system and determine what it represents. Rather, it is the facts about what it would be doing if it were operating according to biological norms. (Millikan 1993c, 10-11)

Millikan seems to be saying that when our cognitive systems function properly, they may well test for the reference-grounding relations identified by various, leading theories of content. However, these other theories don't capture the nature of reference; to truly understand reference, we must see that humans only refer because they are *supposed to*, as a biological imperative. A fundamental understanding of representation is thus achieved only by understanding the role of proper functions and Normal conditions in determining representational content, according to Millikan.[16]

Millikan overstates (and, I dare say, misrepresents) the role of biological function in fixing reference. It is not simply that the cognitive system exploits certain causal connections in order to represent Normally. The biological norm itself, on which intentionality (and thus, representation) is based, is the norm of testing for the presence of a certain causal relation. It is, presumably, the value of detecting this causal relation that makes it worth our cognitive system's while to bother checking for consistency at all: it is the exploitation of this causal relation that explains why testing for consistency is to the advantage of organisms who can so test. Thus, a theory such as Dretske's informational theory of content is not merely tangential to Millikan's biologically-based theory of content. For the naturalistically-minded philosopher, a theory such as Dretske's (or one such as Fodor's asymmetric dependence theory; see Fodor 1987, 1990) is essential to solving the puzzle of intentionality. As a theory of intentional content for LOT terms, Millikan's theory tells us that it is our biological function to refer without informing us of the nature of reference, and this undermines her semantic theory for natural language. I conclude, then, that those who wish to understand what natural relation constitutes reference must look elsewhere.

References

Davidson, D. (1984) *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Davies, P. S. (1994) "Troubles for Direct Proper Functions", *Nous*, 28, 363-381.

Dowty, D. R., Wall, R. E., and Peters, S. (1981) *Introduction to Montague Semantics*. Dordrecht, Holland: D. Reidel.

Dretske, F. (1981) *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Dretske, F. (1988) *Explaining Behavior*. Cambridge, MA: MIT Press.

Elder, C. (1994) "Proper Functions Defended". *Analysis*, 54.3, 167-171.

Fodor, J. (1975) *The Language of Thought*. Cambridge, MA: Harvard University Press.

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

Fodor, J. (1990) *A Theory of Content*, Cambridge, MA: MIT Press.

Fodor, J., and LePore, E. (1992) *Holism: A Shopper's Guide*. Oxford: Basil Blackwell.

Fodor, J., and Pylyshyn, Z. (1988) "Connectionism and Cognitive Architecture: A Critical Analysis". *Cognition* 28, 3-71.

Lakoff, G. (1987) *Women, Fire, and Dangerous Things*. Chicago, IL: University of Chicago Press.

Maloney, J. C. (1994) 'Content: Covariation, Control, and Contingency'. *Synthese* 100, 241-90.

Millikan, R. G. (1984) *Language Thought and other Biological Categories*. Cambridge, MA: MIT Press.

Millikan, R. G. (1993a) 'On Mentalese Orthography, Part 1'. In Millikan 1993c, 103-21.

Millikan, R. G. (1993b) 'On Mentalese Orthography'. In *Dennett and His Critics*, B. Dahlbom (ed.), Oxford: Blackwell, 97-123.

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

NOTES

* I would like to thank Charles Chastain, Walter Edelberg, and anonymous referees for their comments on earlier drafts of this paper.

[1] All page references are to Millikan 1984 unless otherwise noted.

[2] For a detailed hypothesis regarding the nature of LOT and arguments in support of its existence, see Fodor 1975, 31-32, Fodor 1987, Appendix, and Fodor and Pylyshyn 1988, 12-14.

[3] Attempts to develop naturalistic theories of mental content that focus largely on the assignment of content to terms can be found in Dretske 1981, 1988, Fodor 1987, 1990, and Maloney 1994.

[4] For arguments to this effect, see Fodor and Pylyshyn 1988 and Fodor and Lepore 1992. Viewing sentence content as a function of the content of constituent terms is a longstanding tradition in the philosophy of language. This approach is formally embodied in Frege's categorial grammar; for an example of the contemporary development of such an approach, see Dowty, Wall, and Peters 1981. For objections to the formal, constituent-based approach, see Lakoff 1987; Donald Davidson's focus on sentences as the fundamental units to which an interpreter assigns meanings also provides a significant contrast to the constituent-based approach (Davidson 1984).

[5] For concerns about the viability of Millikan's definition of a reproductively established family, see Davies 1994. For a defense of Millikan's definition, see Elder 1994.

[6] Fodor claims that Millikan's emphasis on the function of particular desires results from a failure on Millikan's part to distinguish between the evolutionary function of a mechanism and the evolutionary function of individuals produced by the mechanism. Having hair has an evolutionary function, says Fodor (1990, 66), but no individual hair has that function. Fodor has chosen a poor analogy. There is some plausibility to the claim that our mechanism for having desires was chosen for because of the successes of individual desires (e.g., the desire to get out of the

way of the charging rhino), or short, connected series of desires (e.g., the desire to get out of the way of the charging rhino, together with the desire to run to the right, together with the desire to jump into the small nook between the two big rocks). In contrast, it was not on account of having any individual hair that any particular ancestor lived to the age of reproduction. The various hairs had to work together as one body. An individual hair does no evolutionary good to the possessor of a hair-making mechanism. However, an individual desire can make the difference between life and death. Insofar as a desire-making mechanism was evolutionarily chosen for, it may well have been chosen for one life-saving desire at a time. Therefore, it seems that certain, individual desires could well have had an important function in common that accounts for the evolutionary choice of the desire-making mechanism. And plausibly, this function is as Millikan says it is, i.e., to bring about certain states of affairs.

[7] Throughout the remainder of the paper, I focus primarily on extensions, ignoring the important distinction between a term's actual extension and its intentional (referential, extensional) content, where the latter is thought of as the content a term has in virtue of the *type* of thing it refers to. In the passage quoted above, Millikan identifies the extension of a term as primary to intentional content in the following sense. An actual extension is something a term has only in the context of the term's appearance in a *specific* true sentence (although for many terms this extension will be the same from one appearance in a true sentence to the next). These actual correspondences then help determine what it is, generally speaking, that tokens of the term, appearing in any context, are supposed to refer to: the correspondences between true sentences and aspects of the world determine something much like the term's intentional (referential, extensional) content.

[8] An anonymous referee suggested this way of explaining the structure of Millikan's theory.

[9] I am simplifying matters here by ignoring the complicated nature of calculating mapping rules based, in the typical case, on numerous past utterances, together with numerous past utterances of transforms of the sentence in question. This simplification does not affect the point I'm attempting to make; it merely makes the point easier to get across.

[10] Appearances may be deceiving here for two reasons: on a practical level, it may not be so easy after all to do the naturalist's dirty work and follow the bees around to see where they're headed and where they wind up; and there may be a degree of indeterminacy metaphysically as to whether the watching bees fly in the direction of a nectar source--almost any direction might be the direction of a nectar source if one heads in that direction long enough. I'll put such worries aside; at the very least, it seems a fairly determinate matter whether or not the watching bees eventually *reach* a nectar source after they've witnessed the dance. Perhaps this is the basis of our feeling that matters are cut in dried when it comes to the performance of the proper function of the interpreting device in the case of the bee dance.

[11] We might attempt to make tighter the parallel between the cases of the bee dance and indicative sentences by attributing to the bees beliefs that were caused by watching the dance and that guide the bees' flight toward the nectar source. The disanalogy I emphasize in the text would remain, however. If bees develop bee beliefs upon watching a bee dance it will be easy to determine whether these beliefs are true, once we've identified 'getting to the nectar source' as the apparent function of the bees' belief-forming mechanisms. Whether the bees' beliefs are true or not is a seemingly simple matter: it depends on whether or not the bees flew off toward or reached a nectar source. In contrast, we have no neat and tidy measure of success in the human case: we have no straightforward, task-specific purpose of forming true beliefs in humans, the successful carrying out of which can be taken as the benchmark of true belief in humans. Human beliefs are too rich and varied to think that the analogy to bee beliefs can be carried so far. My inclination is to leave out any talk of bee 'beliefs', and Millikan seems to have similar inclinations; at the very least she makes it quite clear that bee dances aren't representations (71, 96, 1993b, 106-7). And given the reasons why she thinks that the dances aren't representations, it is very unlikely that she will want to include any talk about bee beliefs in an explanation of why bee dances have intentional content (see note 2 to chapter 6).

[12] I have characterized the problem as one of circularity: the consistency tester performs its proper function only if it creates true beliefs; true beliefs can exist only if mapping rules are in place; but mapping rules can only be in

place if the consistency tester has performed its proper function. Looking at matters historically, one might prefer to characterize the problem as a regress, the complaint being that Millikan's theory provides us with no appropriate tools for ending the regress. The choice of characterization matters little to my main argument, however.

[13] As noted in section II.B, Millikan might say that natural language sentences have their content determined independently of belief content, but then she is left without an account of how content is fixed for indicative sentences in natural language.

[14] P1 does not entirely capture Millikan's theory, for it does not address the issue of novel sentences. In the typical case, the content of a novel sentence is determined by the adaptation of the sentence type, e.g., indicative, to the terms in the novel sentence (97-98). For this to occur, the terms appearing in the novel sentence must themselves already have content of a sort (whatever it is that they are supposed to map onto). This can be the case, however, only if the contents of the terms that make up the novel sentence were previously fixed as a result of their appearance in true sentences whose content was fixed by P1 (or if the terms in the novel sentence have appeared in true sentences whose constituent terms previously appeared in true sentences whose content has been fixed by P1, and so on--wherever the recursive iteration stops, P1 fixes content). Thus it is fair to think of P1 as Millikan's fundamental, content-determining principle.

[15] Dretske presents his information-based theory of content in Dretske 1981. He details his more recent views on these matters in Dretske 1988. It is in the latter work that Dretske uses the term 'indicator' and its cognates. However, the concept of indication is very similar to that of carrying information (see Dretske 1988, 58-59, where Dretske remarks on the closeness of the connection between the two ideas), and for present purposes, lumping together the two types of theories does no harm.

[16] Note that Millikan sometimes takes a stronger stand against covariational or informational theories as theories of intentional content, characterizing these alternative views as, not merely beside the point, but generally false:

"Intentional icons do not, as such or in general, carry 'natural information' [with an endnote note here identifying 'natural information' as information in Dretske's sense]. Nor do they 'covary' with or 'track' what they icon." (1993b, 100) Here Millikan is concerned about the gap between Normal conditions and statistical trends, which trends may mislead us if we make them the basis of our content assignments.