Millikan's Contribution to Materialist Philosophy of Mind

Nicholas Shea

Matière Première, 1, Feb 2006, pp. 127-156.

Abstract

One of the great outstanding problems in materialist philosophy of mind is the problem of how there can be space in the material world for intentionality. In the 1980s Ruth Millikan formulated a detailed theory according to which representations are physical particulars and their contents are complex relational properties of those particulars which can be specified in terms of respectable properties drawn from the natural sciences. In particular, she relied on the biological concept of the function of a trait, and the existence of historical conditions which enter into an evolutionary explanation of the operation of that trait. The present article is an introduction to this influential theory of intentionality.

(1) Introduction

The behaviour of humans and other animals is often explained by attributing to them internal representations on which they act. This is a commonplace both of everyday life and scientific psychology. However, neither philosophy nor psychology yet has a good account of the content of these representations. That is one of the most important foundational questions in the philosophy of mind: what makes it the case that mental states represent as they do? The project of 'naturalising intentionality' seeks an answer to that question which is applicable within the sciences of brain and behaviour. The aim is to make representational contents scientifically acceptable by showing how they are related to the other natural properties found within those sciences. Intentional properties are unlikely to be reducible to the properties of basic physics, but an adequate theory of representational content should show how there can be space in the natural world for representation.

This puzzle about the nature of intentionality is one of the two great outstanding

problems in materialist philosophy of mind (the other being to give a materialist theory of consciousness). Like consciousness, as soon as we start to think about the phenomenon of intentionality it begins to seem decidedly odd. How can a representation be "about" anything at all, in the absence of a rational agent (endowed with intentionality) who takes it as being about, or standing in for, some other object or property? An explanation which presupposed the existence of a rational agent to interpret every representation would just postpone the mystery, because it would give no account of the intentionality of that agent's own mental states. Thus, after only a little reflection it is apparent that our commonsense understanding of a representation's aboutness is regressively homuncular if it is applied to mental representation itself. In what, then, does this aboutness consist?

The first step towards a solution is to understand representations as physical particulars. This was achieved by Chomsky's cognitive revolution and promulgated within philosophy by Jerry Fodor (Fodor 1975, 1987). As a physical particular, a representation is part of the material world and can play a role in the causal order. It's content is just one of its many properties. To specify that property may involve mentioning the objects and properties which the representation is about. For example, in telling you that Pierre believes that dogs bark, I have picked out his mental state by its content (that dogs bark), and by telling you what type of cognitive attitude Pierre has towards that content (believing, as opposed to, eg, hoping that dogs bark or guessing that dogs bark). So I have mentioned some external objects and properties to specify the content of Pierre's belief. I could also pick it out in different ways: as the occurrent belief that was most salient to Pierre at 23.00 on 1.1.05 or, if the details of the physical realisation of the vehicles of thought in human brains were known, I could pick it out under that kind of description (as the pattern of coherent synchronous neural firing at x Hz distributed across such as such cortical areas in Pierre's brain, or whatever). Since the contentful way of picking out the belief mentions some things external to Pierre, his belief's having that content could perhaps be a matter of it being in an appropriate relation to those things. Compare the follow property which applies to some parts of Paris: *being on the left bank of the Sienne*. That property applies to regions which could equally well be picked out without mentioning the Sienne (eg, latitude and longitude). It is a relational property of some parts of Paris, and the property is given by mentioning the thing (the Sienne) to which those districts are related. Going back to mental states, when a racket wakens Pierre at 23.00 on 1.1.05 and brings to mind his belief that dogs bark, that belief is related to some external things (a dog, its barking): the belief has the relational property of being caused by the dog's barking. But belief's content cannot be some simple relational property, in the way that *being on the left bank of the Sienne* is a simple relational property of parts of Paris. Why not? Because in the case of beliefs, a belief can still have the content that dogs bark even in the absence of the putative relata (dogs, barking). The belief that dogs bark can occur to Pierre in the absence of any currently barking dogs: later, perhaps when he is remembering the miserable night of 1.1.05, or in a different context, for example when

reasoning about animals and their noises. Compare the simple relational property of *being on the left bank of the Sienne*. Take away the Sienne and no part of Paris would have that property. Not so for beliefs and their contents. So the content of a belief cannot be some simple synchronic relational property of a particular mental state.

Instead, the idea has grown that a representation's content may be specified by some complex relational property of the representation.[1]. If the complex relational property that is intentionality can be fully specified in terms of properties and relations that are found in the natural sciences, then this approach will both show how there is space in the natural world for the phenomena of intentionality, and it will provide explanatory connections between the intentional and the natural sciences. So the problem of naturalising intentionality would be solved if someone could only show how the aboutness of a representation was just a complex relational property of a physical particular.

However, it must be a pretty special relational property, since it founds a distinction between representation and misrepresentation, and perhaps even between truth and falsity. If representations are just physical particulars in the world, how can they be incorrect? *Prime facie* objects aren't correct *or* incorrect, they just are. Consider again poor Pierre's bleary-eyed state when he was woken at 23.00 on 1.1.05. Suppose, in his confusion, he had thought to himself *cats bark*. As a physical particular, this is just another thing in Pierre's head. We can still pick it out as the occurrent belief that was most salient to Pierre at 23.00 on 1.1.05 or as the pattern of coherent synchronous neural firing at x Hz distributed across such as such cortical areas in Pierre's brain (or whatever). This time, it has some different complex relational property which is its content, one that we specify by mentioning cats, amongst other things. However, there is another crucial difference between the two cases: in one Pierre is thinking something *true*, in the other something *false*. And the truth or falsity of Pierre's belief seems to be fixed by its content, together with some facts about the world (in particular, facts about the nature of canine and feline vocalisations). Belief contents must be some complex relational property that is suited to forming the basis of this distinction between truth and falsity. Arguably, not any old descriptive distinction will do here, since it is a difference that can have normative import. There is a sense in which a representation that is misrepresenting is doing something wrong. What kind of complex relational property can be like that?

Ruth Millikan's great contribution to materialist philosophy of mind was to formulate a detailed theory according to which aboutness is a complex relational property of physical particulars. Furthermore, her theory made it plausible that normative force could be applied to the distinction she drew. She did so by relying on a normative notion in the special science of biology, namely the biological concept of function. Millikan relied on work in the 1970s which showed that the concept of biological function is naturalistically acceptable, so that it is unproblematic to think of a trait's biological function as amongst

[1] That is, the contentful property is either identical to that complex relational property or is metaphysically determined by it.

its ordinary physical properties. The beauty is that functions support a normative notion of malfunction (even if biological functions are not in themselves normative). In the 1980s Millikan showed how biological functions could be deployed in the context of the very special systems that are representational to underpin a naturalistically-acceptable account of representational content (Millikan 1984, 1989).

Millikan's theory of intentionality is not the only philosophical attempt to get over metaphysical scruples about the existence of content in the material world by understanding it as a complex relational property. Another leading contender in the 1980s attempted to understand content in terms of the information carried by a representation (Dretske 1981), where information is a matter of causal correlation between properties of the representation and other properties. However, Millikan's theory got closer to capturing an intuitive idea of the content of representations. It also does a better job of illuminating why content attributions have the explanatory force they do in accounting for an organism's behaviour. Of course, her theory faces difficulties and objections. However, even if it does, in the end, need significant amendment, or even radical revision, it does have the merit of demonstrating that the puzzle about the metaphysical status of intentionality in a material world is soluble – representations and their contents may just be physical particulars and their natural properties, respectively.

In the next section I will set out the structure of Millikan's theory. Section (3) goes into some of the details and section (4) explains why the theory delivers representational contents with the required level of determinacy. Finally, section (5) mentions two objections.

(2) Producer, Representation, Consumer

If intentionality were just a matter of biological function then intentionality would pervade every aspect of the living world. Every living organism is packed-full of traits with biological functions, from the protein coat of a virus particle to the delicate structure of the human eye. Most of these traits lack intentionality. Neither the virus coat nor the human eye is *about* anything. They may function or malfunction, but there is no sense in which they can be false. Intentionality is a feature of only those traits with a very special sort of function: a function involving representation. That gives us an outline: biological functions derive their normativity from natural selection, and intentionality arises in a small subclass of biological functions, namely those traits which have the biological function of representing something. So the remaining task is to explain what it is to represent. Sadly, that question is not a great advance on the initial problem of explaining intentionality. We have succeeded in reformulating the problem in different terms, but we are left with the crucially difficult task of elucidating what is distinctive of representation. The new outlook gives hope, because it suggests that the right approach is to look for

intentionality in nature, and then to discern what is distinctive of it, considered in the light of biological functions. If representation can be adequately characterised as a biological function then Darwinism will provide a source of normativity. Otherwise, natural selection just remains a promising analogy.

To address such a difficult task it helps to start with the easiest possible examples. So Millikan looked for intentionality in simpler cases than human cognition. Nature provides a rich source of examples. For, although not every evolved trait has intentionality, there are lots of traits in other animals that look as though they might. The strategy is to find a class of natural traits that are plausibly intentional, and to use those examples to determine what distinguishes intentionality from other biological functions. Animal signalling furnishes good examples which are simple enough to see what is going on. For instance, beavers signal danger by splashing their tail on the water. Other beavers respond to this splash by diving underwater, often returning to their lodge via its protected underwater entrance. This splashing behaviour has a biological function – to avoid predators. But the splash itself also seems to be about something: danger. If it is produced when there is no danger then something has gone wrong: the splash says there is danger when there isn't, so it says something false.

Other animals have much more complex signalling systems. Honeybees are a favourite Millikan example. Bee dances have become one of her trademarks. In a photograph of Millikan in front of the giant honeybee exhibit at the New York Science Museum she seems to acknowledge this with an ironic smile. It is now well-known that bees perform a complex dance inside the hive to signal to other bees the location of a source of nectar. That was originally a startling discovery. At first people refused to believe that one bee could 'tell' another where to find honey, 'representing' in a dance the direction and distance of groups of flowers. But the evidence built up. Bees returning to the hive don't dance at random. Rather, they waggle to and fro along a straight line. The direction of that line corresponds systematically with the direction of the flowers they have just visited. In addition, the number of waggles they make along the line corresponds with how long it takes to get to the flowers, and the vigour of the dance varies with the desirability of the food source. These correlations are already impressive, but the knockout finding is that bees watching a dance behave in a systematic way (since it is dark inside the hive, the 'watching' is done entirely by touch). They fly in the direction indicated by the main axis, and keep flying for the amount of time indicated by the number of waggles they observe (this description is slightly simplified, but the details do not affect the validity of the example). These observations together provide compelling evidence that bee dances do indeed signal the location of nectar. For Millikan this discovery was an important addition to her armoury of examples, because it acts as an intermediary between the blind purposes of biofunctions and the complexities of representation. On the one hand, bee dances do seem to be representational. On the other hand, bees are relatively simple creatures, making it unquestionable that their traits can be explained in terms of evolution

by natural selection. Philosophers might suspect that something special and *sui generis* is needed to explain the intentionality of thought, but no such case could be made for the bees' representations.

Millikan's task was to discern what constitutes the difference between intentionality and other kinds of natural functions. The two examples of the beaver splash and the bee dance illustrate the features that underpin this distinction. They are useful examples because the representations are out in the open, produced by one organism and used by another. This structure is not so apparent when considering internal representations. Millikan's insight was to observe that representational systems are always divided up into a representation, a system that produces that representation, and a system that uses that representation in order to perform some further function. In bees, the dances are the representations. Bees have a system that uses those representations to fly to flowers. Millikan calls the system that makes the representations the 'producer', and the system that makes use of them the 'consumer'. In bees the consuming system sees the dance of another bee and ensures the watching bee flies a certain distance in a given direction as a result. The function of the whole representational system is to take the bees to nectar. That is divided into two sub-functions. The producer's function is to produce a dance that corresponds to the location of nectar. The consumer's function is to take the bee to the location indicated by the dance. A particular dance mediates between these two systems. It acts as a representation that signals to the consuming system the location of nectar, as indicated by the producing system.

The same articulation into a producing system and a consuming system can be seen in the beaver's splashing behaviour. The function of the whole set-up is to keep beavers away from danger. The consumer mechanism contributes by taking the beaver to safety when it hears a splash. The producer mechanism contributes by making a splash when danger is observed. The splash itself sits in between producer and consumer, acting as a representation of danger. To apply this thinking to mental representation Millikan will divide the internal processes involving representations into a system that produces representations and a co-operating system that consumes them. But, unlike the beaver and honeybee cases, both these systems will be inside the same thinker, so that the intermediate representations will also be internal.

Do these natural representations, the bee dance and the beaver splash, really admit of misrepresentation? A representational system can go wrong in several ways. Something might fail in the consuming system. For example, a beaver might hear a splash but fail to take cover. Alternatively, there may be something wrong with a given representation. Think of a splash that is produced in the absence of danger. The splash represents danger and the beavers dive, but there is in fact no danger around. In that case, Millikan argues, the splash is false: it means *danger here now,* but there is none. The same applies to the bees. A consumer bee can fail to follow the directions in the dance it has seen. It could be blown off-course, or run out of energy before it arrives. However, the dance production-

consumption system can also fail to lead the consumer bee to nectar for a different reason: because the dance represents the wrong location. Suppose the dancing bee makes the wrong number of waggles, or gets disoriented and dances in the wrong direction. Then there will be no nectar in the direction indicated. The dance represents *nectar at distance r in direction* θ when there is none at (r,θ). Millikan argues that in such a case the dance is literally false.

So Millikan saw that representing is not a function of a single undifferentiated system. Instead, representation takes place in a system composed of a producer and a consumer, with representations as intermediates produced by one and used by the other in order that the entire system should perform in the way natural selection designed it to. In the bee and the beaver the producer and consumer systems operate in different organisms, but more typically the systems will lie within the same organism and the intermediate representations will be internal. The same distinction applies in both cases. This distinction allows her to focus on the important role of the consuming mechanism in determining the content of the intermediate representations – what I will call a 'forward-looking' approach to content.

Millikan's approach to intentionality relies on the progress that was made in the 1970s in understanding the biological functions. On the face of it, it is extremely odd to explain the nature of some trait in terms of its function (eg, that the opposing thumb is for grabbing): how can the effect that some trait has (eg, manipulating fruit) explain its structure? To explain a cause by its effects gets things exactly the wrong way round. The insight developed in the 1970s was to see that these cases of natural teleology are unobjectionable in the context of evolution by natural selection. The effects of past instances of the trait (eg, opposing thumbs in human ancestors) occur earlier in time than the current trait whose structure is to be explained, and these ancestral effects do enter into a causal story of why there are instances of this trait around today (ancestral grabbings of fruit are part of the reason that there are organisms with opposing thumbs around today). Since natural selection cares only about results, biofunctions must depend ultimately upon a system's effects, not upon what causes it to behave as it does. Individuating a trait by its function, in these kinds of cases, turns out to be a matter of individuating it by its (ancestral) effects.

Others theorists of intentionality have tried to make progress with the idea that it is a function of a representation to indicate, covary with, or carry information about some external state of affairs (Fodor 1990). This idea cannot stand alone, since looking at the circumstances in which representations are produced does not connect with their effects. Millikan's forward-looking theory surmounts this problem by explaining what it is to have the function of detecting or indicating some external state of affairs in terms of the functions of the consumer mechanism.

The importance of the forward-looking approach can be seen in some real-life experimental research. The visual system of the frog was studied in the 1950s. Ganglion

cells were discovered in the retina, very early in visual processing, that respond selectively to little black dots moving across the frog's visual field. Lots of objects and patterns were presented to the frog's eye, but this particular cell only fires strongly when a black pellet (say, on the end of a thin piece of wire) is moved in front of the eye. None of the other patterns caused this cell to respond. It was surprising to discover such responsiveness to a higher-level feature of the scene presented, rather than a basic visual attribute. There is then a question about the function of these cells. Are they merely carrying correlational information about movement which is used in later processing, or are they themselves representing something? The question was answered when it was observed that these cells trigger the frog's fly-catching reflex. Activity in these cells is enough on its own to trigger the mechanism by which the frog darts its tongue out in the direction of the moving pellet. This reflex is stimulated even if the cells are artificially activated by applying an electric current. The tongue reflex is produced as surely as, in 18th century experiments with electricity, applying an electric current to a frog's leg made it contract. So researchers were able to conclude that these ganglion cells represent the presence of a fly (or more precisely, some flying suitable prey). That is a commonsense conclusion. However, notice that it relies on the strategy that Millikan subsequently made explicit: to look to the consumption mechanism in determining the content of a representation. And this is a case, unlike the bee dance and the beaver splash, the representation is an internal state mediating between an internal producer-subsystem and an internal consumer-subsystem. The 1950s researchers adopted the forward-looking approach to work out what these internal cell-firings meant. Millikan's theory endorses this forward-looking approach generally, both for external representations like the bee dance and beaver splash, and for internal representations like the frog's fly-detector cells. In the internal case the theory requires some internal subsystem that, in order to perform its function, consumes the representations.

Producer	*Representation*	*Consumer*
Beaver which splashes its tail on detecting danger	Tail splash	Other beavers: dive underwater
Bee which dances on returning from nectar	Bee's dance	Other bees: fly in the direction & for the distance indicated
Frog's eye and optical system	Firing in a retinal ganglion cell	Tongue-dart mechanism in the same frog

Fig. 1 Producer, Representation and Consumer

In summary, the content of a representation depends crucially upon what the system

consuming it takes it to represent. But beware. That cannot be a final answer, since it still makes an unexplained use of 'represent', which is the phenomenon we set out to explain in naturalistic terms. We have seen the problems with talking loosely: the content of a representation is X if the function of that representation is *to represent* X. How do we adapt this in the light of the idea that we must look forward, to consumers, as determinative of content? Here is a first try: the content of a representation is X if the consumer mechanism (when well-functioning) takes it to represent X. But then we still have the unexplained use of 'represent'. How to eliminate it? Millikan had an insight about that too. Her idea is that what a representation stands for cannot be found in the function of any system, but instead in the conditions in the world that must be in place if the system is to function as designed.

Consider again the bee dance. It is performed on a vertical wall inside the hive. The angle of the main axis of the dance from vertical is interpreted by the consumer mechanism (in other bees) as representing the angle between the direction of the food source and the sun. The observers are caused to travel in that direction, and to travel a distance that is a mathematical function of the number of waggles observed. Suppose the bee dances *n* waggles in a line at angle θ to the vertical. The mathematical transformation performed by the consuming system causes the bee to fly *r* metres in direction θ from the sun. Provided the consumer system calculates the mathematical transformation as it was designed to, the consumer system will be well-functioning. That is, it will be functioning in the same way as accounted for the past successful operation of the mechanism (which in turn led to reproduction of more bees, and hence of more copies of the mechanism). However, well-functioning does not guarantee success. Success requires that there really is nectar *r* metres away in direction θ. That is a condition for the successful performance of the mechanism's function. The consumption system can only succeed in functioning as designed if there is nectar at (r, θ). This condition for the normal performance of the consumer's function is the same as the content of the representation: that *there is nectar at* (r, θ). Therefore, Millikan says (roughly) that the content of a representation is found in the normal conditions for the performance of the function of the mechanism that consumes it. Amongst those conditions are many common background requirements, such as sufficient nutrition, being at the right temperature, etc. But some condition is specific to the consumer's function in relation to a specific representation. That condition is the content of that representation.

This idea also works for internal representations. Many bacteria in lakes and oceans display magnetotaxis: they swim in the direction of lines of magnetic flux. In the northern hemisphere they swim towards magnetic north. This is achieved by means of microscopic particles of a magnetic mineral enclosed within a membrane inside the bacterium, called a magnetosome. The local magnetic field acts on the magnetosome and rotates the whole bacterium towards magnetic north. The bacterium's constant forward-swimming motions

then take it in that direction. In the northern hemisphere, the magnetic field not only points towards northwards, but also broadly downwards, since lines of magnetic flux descend into the earth. (Compasses are designed to compensate for this downwards force, so that the needle points north horizontally – which is why northern-hemisphere compasses perform badly south of the equator.) The effect of this simple connection between magnetosome and swimming is that the bacteria stay away from the oxygen-rich water at the surface of the sea, which is toxic. The swimming mechanism was designed to function by taking the organism in the direction indicated by the magnetosome. However, it is a condition for the normal performance of that function that that is the direction of oxygen-depleted water. That is, the magnetosome-to-swimming mechanism functioned successfully in the past, causing the reproduction of the bacterium and hence of the mechanism itself, only when the magnetosome pointed in the direction of oxygen-depleted water. That is the condition for successful performance of its function. Hence, the orientation of the magnetosome represents the direction of oxygen-depleted water. That is what it is *about*. And when the magnetosome points at oxygen-rich water (deflected, say, by a passing ship) it does something *wrong* and says something *false*. So we have aboutness, normativity and the capacity for misrepresentation. In short: intentionality.

(3) Biological functions and evolutionary conditions for their successful performance

We've seen that Millikan's starting point is to think about functions in the evolutionary sense. That involves looking at what a trait did in the historical past. Millikan came independently to the idea of characterising biological functions historically (aetiologically), although several philosophers had expressed this view in the 1970s. The idea is that some of the things a trait has done in the past will have contributed to the survival and reproduction of the organism carrying that trait. Consider moths' patterned wings as an example. A pattern that matches the background in a moth's environment will help that moth to survive until it can reproduce. The patterns on moth wings are inherited by offspring, so the moth's offspring will have the same pattern, which will in turn help them to survive and reproduce. Now consider a particular pattern on the wings of an individual forest-dwelling moth in the present day – a dappled motif, say. Part of the explanation of why that wing is dappled is that dappled wings in its ancestors helped them to survive and reproduce. The current pattern is caused by genes copied from moth to moth in the ancestral chain leading up to the present day. The ancestral dapples are part of a causal explanation of this copying process, and hence of why we see the individual's dappled wing today. That is the idea behind Millikan's biofunctions. In ancestors a trait did something (call it 'X') that contributed to that trait being copied into the next generation, so that a causal explanation of why the trait exists now must advert in part to the trait doing X in the past – in that case X is one of the evolutionary functions of the trait. Applied to the moth

case, camouflage is an evolutionary function of the dappled pattern on the moth's wing.

Next, think about what happened to all the moths in the past that had this camouflaged wing pattern (those that died as well as those that survived to reproduce). Some will have drifted into unsuitable environments or landed on backgrounds that did not match their wing pattern. In them, the wing pattern will not have aided survival. In others, for all the cunning of the pattern, a perceptive predator may have spotted them anyway, and gobbled them up. To function as camouflage a wing pattern need not be effective all the time, or even most of the time. All that is needed is that camouflage sometimes helped a moth avoid getting caught when without the camouflage it would have been. If there are some such occasions, then the wing pattern in one generation partly explains its being copied into the next generation. And if there are enough such occasions, then natural selection will explain why the moth population evolved so that they all eventually carry the camouflaged wing pattern (i.e., the ones that do not are selected against). Now focus on the occasions when the dappled wing pattern performed its evolutionary function in the past. Those occasions share a crucial characteristic: the moth must have been located against a dappled background. It is a truism that camouflage only works against the appropriate type of background. (Arctic hares stand out as nice white blobs in Cotswold Wildlife Park in England.) So it is a condition for the dappled wing pattern to perform its evolutionary function that the background be dappled, at least roughly. Only when the environment co-operates in this way can the evolutionary function be performed. Being on a dappled background is an environmental condition needed if the pattern (the trait) is to perform its evolutionary function. In general, there is some set of environmental conditions which are necessary to the performance of any given evolutionary function. These evolutionarily suitable conditions include lots of stable background material, like having sufficient nutrition, the surroundings not being too hot or cold, and so on. But they also include conditions peculiar to a particular trait, like being on a dappled background. All of these conditions, the general and the specific, are the evolutionary conditions for performance of that evolutionary function.

Those are the two key biological ideas behind Millikan's theory. The other parts of her theory are more specifically concerned with intentionality, although they rely on examples in the natural world too. To recap, the two ideas are (1) that organisms and their parts have evolutionary functions, fixed by what those mechanisms did in their direct ancestors in the historical past; and (2) there are evolutionary conditions needed for the successful performance of those evolutionary functions. As a signpost to her writings note that Millikan refers to (1) as 'proper' functions and (2) as 'normal' conditions ('Normal' in LTOBC). However those terms have proven misleading. Commentators often forget that Millikan gives the terms technical meanings which do not fit their normal interpretation. The label 'proper' makes it seem as if Millikan's theory starts with an assumption that traits have functions which it is proper for them to perform – that they are supposed to perform – whereas Millikan's theory actually explains rather than assumes the existence of natural

purposes. The term 'normal' is even more problematic, causing critics to forget that evolutionary functions may not be performed very often and that evolutionary conditions may be far from statistically normal – a fact emphasised by Millikan. Indeed, in some cases the evolutionary conditions for performance of an evolutionary function arise only very rarely (like the traits of plants in the Atacama desert designed only for when it rains, which it does on average once every hundred years).

Recall from the last section that the first move towards using evolutionary functions to account for intentionality is to divide representational systems into a producer system, a consumer system, and intermediate representations. For this to be a legitimate theoretical step there must be some principled way of making those distinctions without relying on the concepts of intentionality or representation. That is, the theory must be able to individuate the vehicles of representational content – the representations – before it goes on to tell a story about the content of those representations. Millikan's non-semantic means of individuating the vehicles of content proceeds as follows. First, find a mechanism which receives a variety of inputs and which does different things on different occasions, depending upon the input. That mechanism is then a candidate to be a consumer mechanism, and the variable items are candidate representations. Now look at the evolutionary function of this putative consumer mechanism, and consider the evolutionary conditions for its operation. In particular, see if the evolutionary conditions for its operation vary depending upon which variable item it consumes. The idea is that to be a consumer the system must do different things depending upon which representation it consumes, and that those different activities must each have an evolutionary function, each with different evolutionary conditions for its successful performance. In such cases there is a sense in which the consumer system is *assuming* that the world is a particular way – a different way in respect of each representation – in that it performs an activity which requires for its success that the world be that way. In such cases the consumer system makes different assumptions in response to different variable items. Therefore, we can say that the consumer system is *reading* the variable items as telling it how the world is and acting accordingly. Think of a table with the variable items in one column, the response of the consumer system in a second column, and a third column listing the evolutionary conditions specific to that response. Here is a portion of that table for the bee dance:

(1) Variable Items: Dances	(2) Response of the Consumer: Flight	(3) Specific Condition: *that there is nectar …*

2 waggles at 45° from vertical	60 seconds flight at 45° to the direction of the sun	200m away to the south west
3 waggles at 90° from vertical	90 seconds flight at 90° to the direction of the sun	300m away to the west
5 waggles in vertical direction	150 seconds flight in the direction of the sun	500m away to the south
…	…	…

Fig. 2 'Look-up table' for the bee dance (in part)

Columns (1) and (3) of the table list the way that the variable items must correspond to the world if the consumer system's activity is to be successful. For bee dances, the correspondence displays a mathematical function relating waggles to distances and angles to directions. However, the correspondence may be much more simple. Consider the table for the simpler example of the beaver splash:

(1) Variable Item	(2) Response of the Consumer	(3) Specific Evolutionary Condition
tail splash	diving under water	that there is danger nearby now

Fig. 3 Look-up table for the beaver splash

That is the whole of the table. The 'variable' items only display minimal variation: splash or no splash. And there is only a very simple correspondence assumed by the functioning of the consumer mechanism. However, for all representational systems there is a way of listing the specific evolutionary condition for each variable item. That will be a list of what specific evolutionary condition is needed for the successful performance of the activity the consumer mechanism performs in relation to each variable item. Put another way, it is a list of the different assumptions the consumer system makes about the world when it is triggered by the different variable items. If we use the metaphor of the representation telling the consumer system that the world is a certain way, then the table is a list of what each different representation tells the consumer is the case.

Actually, the idea of a list is a simplification. No finite list could capture the relation between bee dances and what they represent since there is an infinite range of possible dances corresponding, for example, to all the possible directions between south and west. What allows consumer systems to respond appropriately to this entire range of

representations is that there is a systematic relation between the dance and the direction represented: a mathematical transformation converts one into the other. Millikan observes that representational systems always display a systematic relationship between what is represented and variable features of the representation, so that the content can be derived by some appropriate mathematical transformation of the variable features of the representation.[2]

Let's recap. How are the different systems – producer, consumer, and system of representations – individuated non-semantically? The answer runs as follows. Look for a mechanism triggered to perform a range of activities by a range of variable items, and if there is a transformation (like the look-up tables above) giving for each variable item a different evolutionary condition for the successful performance of that mechanism's activity, then the mechanism is a consumer system and the variable items are representations. In turn, the mechanism that produces the representations is a producer system. This is to state the theory in terms of a different function of the consumer system in response to each representation, with a different condition for each function. However, we can be more succinct and generalise across all the different outputs of the consumer system, and across their different conditions for successful performance of their functions. There is a condition which covers them all: that the representations should correspond to conditions in the world in the way given by the transformation. That is the evolutionary condition for the successful performance of the evolutionary function of the consumer mechanism.

Although Millikan's theory is forward-looking, relying on the consumer mechanism to fix content, she has similar things to say about the producer mechanism. The evolutionary conditions for the performance of the evolutionary function of the producer mechanism – what was the case when the mechanism did something which actually helped the survival of its ancestors – will include the fact that the representations corresponded to the world according to the entries in the look-up table. Ancestral beaver splashes contributed to survival only when there was danger. So agreeing with the look-up table is not the *function* of the producer system, but it is an evolutionary *condition* for the performance of its evolutionary function. For both producer and consumer systems, it is important to switch from focusing on functions to conditions for successful performance of evolutionary functions.

Now that the theory has a way of individuating representations non-semantically, it still has to say what each individual representation is about. That is, we still need to know how the content of each representation is determined – how to 'pin the content on the representation'. For Millikan's theory, that's easy – use the look-up tables! The content of

[2] That there is some 1-1 mapping between variable features of the representation (ie, between different representation types) and the contents that are represented follows trivially from the way the theory fixes content. In the case of the bee-dance, that function can be specified by a simple mathematical formula; but that is not essential.

a given representation is what the consumer system assumes it to mean. More precisely, a given representation is *about* the specific environmental conditions that were evolutionarily present when the consumer mechanism performed the evolved function of the activity triggered by that particular representation (i.e., the condition specific to the activity triggered by that representation).

(4) Determinacy and the disjunction problem

Those are the bare bones of the theory. We saw in the last section that there was a worry whether teleosemantics could deliver contents with the required determinacy. The answer is that it does. Rival theories of intentionality based in causal covariance or information have a big problem with determinacy. A whole collection of difficulties in this area are referred to in the literature as 'the disjunction problem'. Millikan's theory was formulated with the disjunction problem in mind and so deals with the difficulties as it goes along. The problem is addressed by parts of the theory we have seen already: the forward-looking switch to the function of *consumers* rather than the circumstances in which representations are produced, the reliance on *explanations* of the evolutionary functions of those consumers, the focus on evolutionary *conditions* for the successful performance of those function, and the reference to those conditions *specific* to a given representation in the context of a range of representations consumed by the same mechanism. It is worth briefly spelling out how these aspects of the theory address the disjunction problems and deliver contents with the required determinacy.

A classic version of the disjunction problem applies to the frog's visual system. Its retinal ganglial cells fire in response to passing flies, but also when a little black pellet on the end of a piece of wire is passed in front of the frog's eyes. Could the frog's visual system be representing something like this: *that there is either a fly here now or there is a little black thing here* now? After all, it is a more sensitive detector of the complex property *fly or little black* thing than it is of the simple property *fly*. Causal theories consider the whole list of items which actually or potentially trigger a given representation, and so tend to ascribe contents which are a long disjunction of items on the list (hence the 'disjunction' problem). Teleosemantics rules out these kinds of disjunctions because not all actual or potential triggers count, but only those things the presence of which are part of the evolutionary conditions for the successful performance of the evolutionary function of the mechanism. That rules out *little black pellet* as part of the content, since ingesting those has not contributed to the survival and reproduction of the mechanism.

Another way of posing this problem is to suggest that the frog's neural firing refers not to *flies* at all, but only to *little black things*. After all, the representation is triggered whenever the frog sees a little black thing passing across its visual field. Even if those little black things are lead pellets, the frog will go on darting its tongue at them. So, is the

representation about *flies (flying prey)* or *little black things ("LBTs")*? Opponents of teleosemantics allege that Millikan's theory cannot adjudicate this dispute, and so cannot deliver the required determinacy. The argument is that the following is a good functional explanation of the operation of the system: the frog's optical system detects LBTs, which is beneficial because in the evolutionary environment LBTs correlate with flies. In fact, Millikan's theory has a decisive answer: when it comes to intentionality we should look at the evolutionary conditions for performance of the function of the consumer system. LBTs are not part of the condition for successful performance. When the consumer (tongue dart) system functioned successfully it caught a fly (more precisely, a nutritional flying object). Recall that Millikan's proposal is to consider only the past instances of acting on the representation which actually contributed to the survival and reproduction of the organism (in a historically normal way), and then to look for an environmental condition that enters into an explanation of the success of those historical actions (and which is specific to actions caused by that particular representation). She is not individuating content by evolutionary functions, but in terms of environmental conditions that enter into a causal explanation. In the case of the frog, that explanation will mention the following environmental condition: that there was a flying nutritional object passing when the frog darted out its tongue.[3] That is a condition of its evolutionary functioning. So the tongue dart system is reading the neural firing as representing flies. That is what the representation refers to.

So teleosemantics takes in its stride two basic versions of the disjunction problem: 'fly or pellet', and 'little black things'. Both concern alternative descriptions of objects out in the world. Another version of the disjunction problem asks about items in the causal chain between the object and the perceiver. How is it that teleosemantics delivers contents referring to objects out in the world rather than referring to items in the perceptual chain, like shadows on the retina or light waves in the air? For clarity I will change the example. Field mice have a simple mechanism which causes them to run for cover whenever a small shadow passes nearby. Many shadows are caused benignly, e.g., by falling leaves. But enough shadows are caused by aerial predators to make it worthwhile for evolution to have designed this reflex. Objectors to the theory say that teleosemantics cannot distinguish between *aerial predator* and *shadow* as being the content of the representation which is triggered (considering the representation that mediates between the optical production system and the consumer system that causes avoidance behaviour). The objection mistakenly assumes that the system functions to detect shadows, with the world co-operating to ensure that shadows correlate with predators often enough to be useful. However, in this case again Millikan's theory gives a determinate answer to the

[3] The condition *that there was a little black thing present* might also enter into some kind of causal explanation of how the tongue dart contributed to the survival and reproduction of the ancestral frog, when combined with the hypothesis that a significant proportion of ancestral little black things were nutritious. However, that is a weaker explanation; correlatively, *that there was a flying nutritious object present* is a better candidate as evolutionary explanatory condition than *that there was a little black thing present*.

question. The theory tells us to look to an evolutionary explanation for the survival and reproduction of the field mouse's mechanism. When we focus only on those occasions when the mechanism functioned so as actually to cause its own reproduction into the next generation, a causal explanation of that function will advert to detecting a *predator*. Put another way, the evolutionary condition that connects with successful performance of its evolutionary function is not *shadow*, about which evolution does not care, but *predator*, about which it does.

In other words, the focus on explaining the actual success of the organism's behaviour leads to evolutionary conditions which concern the kinds of things *in the world* which evolution cares about: predators, food, mates, etc. The alternative – to look to the actual discriminatory capacities of the systems – is always a temptation, but one which is avoided by Millikan's teleosemantics.

For completeness I will consider a final version of the disjunction problem. This asks teleosemantics to arbitrate between various things as potential contents which are all evolutionary goods for the system in question. To distinguish from the other disjunction problems I will switch example again. Return to the bacterium, with its magnetosome pointing downwards, to magnetic north. I said that the orientation of the magnetosome represents *direction of oxygen-free water*. But why does it not represent something more general like *safety* or *survival environment*, or even the very general, *something evolutionarily beneficial*. That last is, of course, a condition for the successful performance of the evolutionary function of every evolved mechanism – *ex hypothesi*. This potential indeterminacy is again removed by different aspect of the theory, namely that a consumer system processes a range of representations and produces different behaviours in respect of each. Not all conditions for the successful performance of the functions of these behaviours count as contents, but only those *specific* to a particular representation. So background conditions like *being at an appropriate temperature* are ruled out as potential contents. As are general conditions like *being evolutionarily beneficial*. Millikan develops this idea further: the specific condition is the one which goes with what she calls the 'most proximal normal explanation' of the operation of the mechanism (recall that 'normal' does not mean usual or statistically normal). There is not space here to go into the details, so what I have said is only a partial clarification. But two important points should be noted. Firstly, the 'proximal' is not an arbitrary addition to the theory to deal with problem cases, but instead arises from the intuitive idea that the candidates for content are the conditions specific to a particular representation, in the context of a range of representations used by the same consumer system. Secondly, Millikan's theory will leave some indeterminacy about the content of representations in these cases, and in other examples of primitive representational systems. That level of indeterminacy is a virtue since it is implausible that such basic systems have as tightly-focused contents as do many representations in humans.

For example, the representation in the frog's optical system has a relatively

unfocused content, something like *moving dark nutritious object*. This is nothing like the kinds of exact, focused contents possessed by human propositional attitudes. However, it is a virtue of teleosemantics that it ascribes these kinds of unfocused contents to primitive intentional systems. It is only in higher representational systems that there is sufficient complexity to generate focused contents. Furthermore, using terms like *moving dark nutritious object* to ascribe content to representations in simple systems can be very misleading, since the fly's representation does not itself contain any of the structure that we use to describe it. The frog has no representations corresponding to the components: movement, dark colouring, nutrition, physical object. Rather it has a single undifferentiated representation, which we do our best to describe using the much more focused articulated concepts of human language. But this too is a merit of the theory. Determinacy increases and disjunction decreases in considering more complex representational systems, and it is only with the constituent structure and general-purpose nature of human intentionality that the kinds of highly-focused contents we are used to are achieved.

(5) Two Objections

(i) Inaccessible contents

This objection to teleosemantics concerns our ability to represent what are called 'inaccessible contents'. A caricatured version of the worry goes as follows: Natural selection can only care about what is causally relevant to survival and reproduction. But distant areas of the universe have never been causally relevant to the survival and reproduction of any organism on Earth. So how can humans have representations with contents like *Sirius is 10^{14} kilometres from the Earth* (which we do)?

The simple reply relies on the generality of human cognition. We take concepts that achieve their content in one context, where they are causally relevant, and reuse them in other contexts where they are not. So the concepts of METRE and TEN TIMES are causally relevant in all sorts of everyday contexts. But the application of them together enough times can create inaccessible contents (*ten times ten times … metres*). Those inaccessible contents can be referred to in virtue of the re-using terms which get their content from everyday contexts.

The debate then becomes much more subtle. The nuanced view of the objection is that the right content in the first place for the everyday use of METRE is *100 cm and not causally inaccessible*. The idea is that every content attribution licensed by teleosemantics will contain the qualification 'and not causally inaccessible' – unnoticed, because it is irrelevant in everyday contexts.. So, for example, the honeybee's dance means *nectar at (r, θ) and (r, θ) is not causally inaccessible.* The objector argues that it is a

consequence of relying on natural selection that these kinds of sub-clauses will constrain the content of all representations whenever they are used, so that when they are combined into complexes like TEN TIMES TEN TIMES ... METRES, the content is *ten times ten times … metres and not causally inaccessible*.

To sketch an answer, there are three strands to teleology's response:- (1) human general-purpose cognition re-uses representational resources in a range of contexts, as mentioned above. (2) the conditions for successful performance of a *derived* function can fall outside anything in evolutionary history, provided they lie within the range of the kind of conditions that explain the performance of the evolutionary function of the relational mechanism from which they derive. (3) the properties employed in scientific enquiry have some scale of naturalness, so that a gerrymandered property like *metre and not causally inaccessible* is less suitable to appear in an empirical theory than the commonsense *metre*. As I have argued above, greater determinacy comes with greater representational complexity. The details of Millikan's theory as set out in LTOBC show how the appropriate amounts of determinacy are achieved at the appropriate levels of representational complexity. It takes the complexity of the human representational system to represent inaccessible contents. But the challenge was only to explain how humans can represent inaccessible contents; simpler organisms probably cannot. As you would expect, the details of the debate about inaccessible contents become very intricate, and there is not scope to give a definitive answer here. It is enough to observe that the objection may well be surmountable.

(ii) Swampman and the historical nature of evolutionary functions

The final type of objection we will consider arises from the historical nature of evolutionary functions. It can be stated briefly, but is potentially very serious. The project is to explain intentionality, but remember that the usefulness of intentionality in the first place is that ascribing representational content is a way of predicting and explaining organisms' behaviour. Why should historical functions be any good for that? Surely we should be more interested in how things currently operate. The objection is made graphic by considering 'swampman', a molecular duplicate of a normal living, thinking human, but which has no history, having arisen instead by random chance (no matter how mindbendingly unlikely that would be – this is, after all, philosophy).

This is a deep and important conceptual problem that teleosemantic theorists must grapple with. Millikan has some answers, but the debate is far from closed. A first observation is that evolutionary functions are a good guide to current organisation, since current circumstances are much like those in the evolutionary past. But that invites the question why current functions, tailored to current circumstances, are not an even better guide. Millikan's answer appeals to the generality of the explanations she offers.

She is attempting to explain biological functions in species. It is widely accepted that

these are historical categories and range over classes of things which, though lacking any defining physical similarity in the here and now, share properties in virtue of their history. In coming across a brown bear one is warranted in believing it will have various of the properties of previously experienced brown bears (in particular, that it is carnivorous and might attack). That warrant derives from the fact that the current bear shares a history with the previously encountered members of the species *ursus arctos*. There is no synchronic causal connection between possessing various perceivable properties (large stature, brown coat) and being disposed to aggression. The inference from perceptible properties to aggressiveness is underpinned by the fact that all members of the species are copied from some common ancestor; an inference from, for example, having a brown shaggy coat to aggressiveness would not be warranted if applied to some other animal that is not in the species. Thus, property projections over instances of species do not extend to individuals that are similar in some respects but do not share the appropriate history.

However, it is not clear how far these considerations can take us when considering intentionality. The aim of the swampman thought experiment is to make vivid the possibility that representational explanations depend only on current properties and allow warranted inference between organisms in virtue of some intrinsic property. So opponents of teleosemantics use swampman to suggest that contentful explanations proceed via in the synchronic underlying nature of the entity whose behaviour is to be predicted and explained, and not via its history – content attribution explains behaviour in the same way that the molecular structure of water explains its liquidity, and is importantly different from the way identifying an organism as a member of a particular species allows prediction and explanation of its properties. The latter proceeds via the entity's history and the former does not. It is no answer to these kind of considerations to point out that if contents are historical properties they will not project to swampman. That is the very issue in question.

A second consideration derives from the normativity that can be attached to representational properties. Aetiological functions seem normative too: a trait which fails to perform its aetiological function is failing to perform its purpose. There is a sense in which it is failing to do what it was designed to do. One can grant that traits also have current, synchronic functions, but still resist the idea that any of these is normative. Whether that conclusion can be supported is a substantial question in the philosophy of functions which there is not scope to discuss here. However, one view is that only aetiological functions underpin a normative distinction. Millikan thinks, independently, that normativity is built into content attribution – so a misrepresentation is failing to comply with a norm which follows from the content of the representation being as it is. If that is right, and if it is only aetiological functions that are normative, then it seems likely that only aetiological functions are up to the job of naturalising intentionality. It follows, then, that swampman has no intentional states; and it is easy, then, to see why – having no history, no norms at all can apply to swampman's states. This line of response will not

suffice to convince sceptics, because they will not accept that the norms of intentionality must be founded in the norms of function. It does, however, give a more detailed explanation of Millikan's motivation for withholding intentionality from swampman.

A third line of response is also available, although it is one that Millikan herself would be unwilling to countenance. Perhaps there is more than one way to naturalise intentionality. That is, perhaps there are different underlying phenomena that licence contributions in different cases. In organisms with appropriate evolutionary histories, it may be right that teleosemantics specifies the contentful properties which underpin our ascription of contentful states to them. That need not, however, rule out the possibility of other kinds of content attribution. In particular, it leaves open the possibility of a different kind of content attribution which would be useful in a world full of swampmen. Of course, the pluralist answer can be turned into an objection – why not apply the theory of content for swampmen to the organisms which share a history too? In order to see whether that reply can be met with an adequate rejoinder we would need the details of the various other theories of content in the pluralist's armoury. However, content pluralism raises the interesting possibility that teleosemantics is the right theory of content for organisms that share a history and have aetiological functions (ie, for actual humans) even if it does not capture another kind of content attribution which could be used to make some sense of the behaviour of an ahistorical swampman.

Many philosophers take swampman and the problem of the historical nature of Millikan's content attributions to be a significant objection to embracing teleosemantics. In my view, whether a convincing rebuttal will depend upon some significant further insight into the explanatory role of content attribution and the nature of contentful explanation, such that it becomes clear whether or not historical properties are suitable for explanations of that kind.

(6) Conclusion

Vigorous debate continues about teleosemantics: objections to its fundamental commitments as well as refinement of its details. It is not yet clear in whose favour the disputes will be resolved. Several things are, however, evident. Firstly, Millikan's theory has quickly risen to prominence as one of the most promising ways of naturalising intentionality. Secondly, it has served to reinvigorate a field of inquiry that had reached a conceptual impasse. Thirdly, the details of LTOBC provide plausible answers to many of the objections commonly raised against teleosemantics. Finally, teleosemantics could be improved in its plausibility were more known about the structure and functions of the components of human psychology, which remains an extremely open area of investigation. At the very least, Millikan's theory has succeeded in reinvigorating materialist philosophy of mind by showing that the ultimate source of the aboutness, normativity and capacity for

misrepresentation in all intentional systems, including ours, may just be found in evolutionary functions.

Nicholas Shea
Faculty of Philosophy and Somerville College
University of Oxford

30 May 2005

References

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

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

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

Fodor, J.A. (1990), 'Psychosemantics, Or Where Do Truth Conditions Come From?' in W. Lycan (ed.) *Mind and Cognition: A Reader* (Oxford, Blackwell).

Millikan, R. (1984), *Language, Thought and Other Biological Categories*. Cambridge, Mass: MIT Press.

Millikan, R. (1989), 'Biosemantics', *Journal of Philosophy* 86, pp. 281-297. Repr. in *White Queen Psychology and Other Essays for Alice* 1993. Cambridge, Mass: MIT Press.