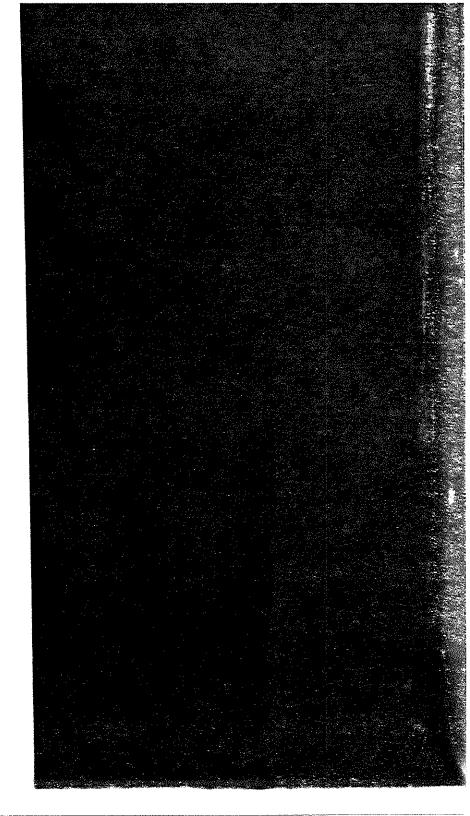# INTENTIONALITY AND TELEOLOGICAL ERROR[1]

BY

PAUL M. PIETROSKI

Theories of content purport to explain, among other things, *in virtue of what* beliefs have the truth conditions they do have. The desire for such a theory has many sources, but prominent among them are two puzzling (and related) facts that are notoriously difficult to explain: beliefs can be *false*, and there are *normative* constraints on the formation of beliefs.[2] If we knew in virtue of what beliefs had truth conditions, we would be better positioned to explain how it is possible for an agent to believe that which is not the case. Moreover, we do not say merely of such an agent that he believes that p when p is not the case. We say the agent made a *mistake*, and often *criticize* him accordingly; we think agents *ought* not have false beliefs, and that such beliefs *should* be changed; *etc*. An adequate theory of content would, presumably, reveal the source of these normative facts about the mental lives of agents. Indeed, it is typically taken to be an adequacy constraint on a theory of content that it help explain the possibility of error and the "normativity" of content. Teleological theories of content promise to do just this.

Teleological accounts are motivated by the following intuition: the content of a belief type B is that property P is instantiated, if B-tokens are *supposed* to be produced when and only when P is instantiated. This accounts for the connection between content and normativity directly. For the idea is that a belief has its truth conditions in virtue of having a particular purpose. So on the assumption that B-tokens are supposed to be produced when and only when P is instantiated, such tokens would be *false* if produced when P is not instantiated. The task facing the theorist, of course, is to say in virtue of what a given belief has its purpose without—on pain of vicious circularity—saying that a belief has the purpose it does in virtue of having certain truth conditions; and here is where appeal to teleology figures in. Common to most teleological

accounts of anything is the idea that systems have parts, and that a part M of a system S is supposed to perform a certain task for S, if S has M *because* (i) M sometimes performs the task, or perhaps (ii) M is a token of a relevant part-type, some tokens of which perform the task. Coupling this idea with the idea that beliefs are supposed to indicate their truth conditions, we get the following Teleological Claim about Content (henceforth, 'TCC'):

Given a system S with a mechanism M that produces tokens of a belief type B in various circumstances, B-tokens have as their content that property P is instantiated, if S has M as a part *because* either M itself or other tokens of the relevant mechanism-type sometimes produce B-tokens when P is instantiated.

As stated, TCC is vague enough that many philosophers may find it an attractive starting point. But the real work involves cashing out the "because" in a plausible way. In this paper, I focus on Ruth Millikan's account of "Proper Functions."[3] Millikan's resulting theory of content is especially attractive. For it is "naturalistic," in that the theory is stated in a non-intentional idiom; and Millikan purports to show how intentional phenomena are continuous with certain non-intentional, but in some sense "purposive," phenomena studied in evolutionary biology. But, I argue, Millikan's account has the following consequence: A system can have the belief that P is instantiated without having *any* systematic ability to tell whether or not P is instantiated (in a given region at a given time). Indeed, instantiations of P can be completely irrelevant to the system's tokenings of the belief that P is instantiated. The corresponding intentional explanations of such a system's behavior will, on Millikan's account, be very implausible; and this gives us strong reason to think that Millikan has not provided an adequate theory of *content*. Millikan's technical apparatus does define a relation that can hold between a system's mental states and properties sometimes instantiated in its environment. But, I argue, the relation so defined is not "has as its content that." Finally, while I cannot address other teleological accounts here in any detail, I suggest that the problems for Millikan's account generalize, in that they give us reason to doubt TCC.

I. *Motivating Teleological Accounts of Content*

I begin with two familiar test cases: fuel gauges and frogs. At least for purposes of exposition, let's say that instances of a fuel gauge needle being in a particular position are tokens of a belief type. With respect to some fuel gauges, at least sometimes, the tank is nearly full when the

needle is near 'F'; and similarly for other needle positions. Suppose, as seems plausible, that my car has a fuel gauge as a part because of such facts. Of course, with some fuel gauges, the float sometimes gets stuck; and so sometimes, the tank is nearly empty when the needle is near 'F'. But *pace* conspiracy theories, my car does not have a fuel gauge because of such facts.

Given TCC, then, instances of my car's fuel gauge needle being near 'F' are tokens of the belief that my car's gas tank is nearly full; and so instances of the needle being near 'F' when the tank is nearly empty are false token beliefs. The same reasoning applies, *mutatis mutandis*, to other needle positions.[4] But my car contains a fuel gauge as a part because some engineer designed the car, and the engineer's intentions were carried out on the assembly line. So one might worry that teleological accounts are relevant only to artifacts that have contentful states (if they do) only because the systems were built by agents who effectively *assign purposes* to parts of the system.

Enter frogs, who have a neural mechanism, in virtue of which they, by and large, snap their tongues when a bug passes through their visual field. But a frog will also snap if an experimenter tosses a dark metal pellet through the frog's visual field. At least for purposes of exposition, let's say that the frog's neural mechanism produces frog-beliefs; and let's grant the intuition, had by some, that these beliefs are true if tokened in the presence of a bug, and false if tokened in the presence of a pellet (and the absence of a bug). TCC delivers just this consequence, if the frog in question has its neural mechanism as a part *because* either the frog's own neural mechanism or other tokens of the neural mechanism-type sometimes produce tokens of the frog-belief in the presence of bugs. The explanation of false belief then takes the following rather attractive form: A system S has a mechanism M that produces tokens of a belief type B when and only when some property Q is instantiated;[5] S has M as a part because either M itself or other tokens of the mechanism-type sometimes produce B-tokens when property P is instantiated; so given TCC, it follows that false B-tokens will be produced whenever Q but not P is instantiated.

Unlike the fuel gauge, however, the frog's mechanism is not the product of any agent's design; and in the absence of a frog-designer or backward causation, properties of the frog's own neural mechanism (including the property of sometimes firing when bugs are present) cannot explain why the frog *has* the mechanism. But the frog's neural mechanism is, presumably, the product of evolution by natural selection; and one can argue that a current frog has its neural mechanism because other tokens of the mechanism-type (viz., those in former frogs) sometimes produced tokens of the frog-belief in the presence of bugs. All of this depends on the assumption that frogs have beliefs. But the frog's

case at least renders plausible the idea that teleological accounts of content can cover non-artifacts. What is required, though, is a theory that makes the relevant "because-claims" for non-artifacts precise, without making illicit appeal to intentional notions.[6]

II. Millikan's Proper Functions

Millikan defines 'ancestor', so that both (a) my grandfather is an ancestor of mine, and (b) my grandfather's heart is an ancestor of my heart; but X is an ancestor of Y, only if there is a (correct) causal explanation of why Y resembles X in certain specifiable respects.[7] (Human genetic structure causally explains both my resemblance, in many respects, to my grandfather and my heart's resemblance, in many respects, to my grandfather's heart.) An item i has the "Proper Function" of performing a task T, if: (1) some ancestors of i performed T; (2) there was some property Q, such that having Q sometimes caused said ancestors to perform T, with the result that having Q was positively correlated with performance of T over some group of individuals including (but not limited to) said ancestors of i; and (3) citing (1) and (2) figures in a correct explanation of why i exists. The explanation mentioned in (3) may directly explain the (re)production of i itself; or it may explain why the items in i's "family"—i.e., the ancestors of i and their descendants, including i—proliferated. The definition is complex, and meant to handle several different kinds of cases; so some examples may be helpful.

Suppose an object found in the woods is used to pound nails. It sometimes gets nails pounded in by virtue of its shape and weight; and because the object is used, there is a positive correlation (over some group of potential tools) between having this shape/weight and nail-pounding. Now suppose a second object, similar in size and weight to the first, is built *because* the first object had the properties just mentioned. This is a straightforward case of (re)producing an artifact that has the Proper Function of pounding nails. Suppose now that certain ancestors of my heart had a complex of structural properties that made them, at least often, successful blood-pumpers; and as a result of their use in former creatures (viz., my ancestors) there was a positive correlation between having said structural properties and pumping blood. The pumping of former hearts doesn't causally explain, at least not directly, why my heart exists. But the pumping of former hearts does explain why organisms with hearts did better at surviving and reproducing than organisms without hearts; and so the pumping of hearts explains, at least in part, why my ancestors and their descendants proliferated. So in that sense, the pumping of former hearts helps explain why I am here; and this explains why my heart is here. Thus, my heart has the Proper Function of pumping blood in the absence of any heart-designer.

Crucially, items can also have "adapted" Proper Functions. Photocopiers have proliferated as a family, because certain ancestor copiers were sometimes able to make something that was "like" (in specifiable respects) whatever was put into their input trays. Current photocopiers thus have the relational Proper Function of making something like (in those specifiable respects) whatever gets put into their input trays. So while photocopiers have not proliferated because they made copies of my (token) paper, the machine down the hall came to have this as its "adapted Proper Function," when I put my original copy of this paper into its input tray. In general: if a mechanism M has the Proper Function of performing a certain operation on *whatever* is such that M bears the relation R to it at time t, then if M bears R to x at t, M has the adapted Proper Function of performing the operation on x at t. If my fuel gauge has the Proper Function of indicating the current level of fuel in my gas tank (whatever it is), then given that my tank is full, my fuel gauge has the adapted Proper Function of registering 'F'. A gauge with a needle stuck at 'F' will thus perform in accordance with its Proper Function whenever the tank is full. But, Millikan says, such a gauge does not perform its Proper Function *in the Normal way*, i.e., the gauge does not perform its task in the way that former fuel gauges performed the task in those historical cases that explain the proliferation of fuel gauges.

Millikan also emphasizes the fact that the Proper Function of one mechanism may be that of helping another mechanism perform its Proper Function. Suppose that when Morty's thermostat is set at '20° C', Morty's heating system acquires the adapted Proper Function of maintaining the temperature in his house at 20°. The heating system will perform this function (in the Normal way) only if it receives "triggers" from the thermostat when and only when the temperature dips below 20°. So the (adapted) Proper Function of the thermostat is to provide triggers in just these circumstances. For it is the fact that thermostats sometimes provide triggers "at the right times" that explains their proliferation. Thus, triggers produced when the temperature is above 20° are imProper; and if the thermostat is not well-calibrated, it may be disposed to produce any number of such imProper tokens. According to Millikan, this just shows that the purpose of items that serve as inputs to a *consumer* mechanism depends on the purpose(s) of the consumer, and not on when the input items are actually produced.[8] This last claim is crucial. For Millikan's central idea is that the mechanisms to which beliefs serve as input have Proper Functions that can be fulfilled (in the Normal way) only if the beliefs are tokened in certain conditions; and these conditions are held to be the truth conditions of the belief. Think of thermostat states as beliefs, and the heating system as the belief-

consumer. When the belief-consumer consumes a belief, it will perform its Proper Function in the Normal way, only if certain external conditions obtain. For given such input, the system will (Normally) blow hot air into the room; and this will serve to maintain room temperature at 20°, only if the room temperature has dipped below 20°.

There are, of course, many necessary conditions for the Proper Functioning of a heating system that has received a trigger from its thermostat: there must be sufficient power, the wiring must be working appropriately, there must be no earthquake, etc. But Millikan thinks she can rule out such "background conditions" as candidates for the content of thermostat-states, as long as the satisfaction of these conditions in certain historical cases does not explain the proliferation of thermostats; and facts about temperature (as opposed to earthquakes) are presumably what explain how thermostats helped heating systems perform their Proper Functions, and hence explain the proliferation of thermostats. This distinction between background and explanatory conditions is neither perfectly sharp nor completely satisfactory. But this is hardly a strong point against Millikan's account, as opposed to other (say, causal) accounts. Let us say, then, that a consumer mechanism C "Needs" that property P be instantiated when it consumes B-tokens if and only if (i) C can perform its Proper Function (in the Normal way) when it consumes B-tokens, only if P is instantiated, *and* (ii) among the necessary conditions picked out by (i), the instantiation of P is what explains why the production of B-tokens in certain circumstances helps M perform its Proper Function (and thus explains why mechanisms that produce B-tokens in those circumstances have proliferated). We can now state Millikan's account of content rather simply: Tokens of a belief type B have as their content that property P is instantiated, if consumers of B-tokens Need that property P be instantiated when they consume B-tokens.

Consider, finally, Millikan's analysis of a case introduced by Dretske.[9] In the ocean, anaerobic bacteria must constantly move towards the bottom and away from the oxygen-rich water near the surface. A certain species of bacteria actually contain tiny magnets that are sensitive to magnetic north. These "magnetosomes" typically propel the bacteria in the survival-appropriate direction. But a magnet in the right place can "fool" the bacteria into oxygen-rich water. Dretske claimed that the magnetosomes serve to represent the direction of magnetic north, on the grounds that such is the information carried by the magnetosome-states. Millikan rejects this claim, because it allows for error only when the magnetosome is, in some sense, "broken;" and she claims that error is typically not a matter of internal malfunction, but a matter of external conditions not being as some mechanism Needs them to be. According to her, the magnetosome represents:

only what its *consumers* require in order to perform *their* tasks . . . What they need is only that the pull be in the direction of oxygen-free water at the time. . . . For that is the only thing . . . the absence of which would disrupt the functions of those mechanisms that rely on the magnetosome for guidance.[10]

We have a mechanism C, whose Proper Function Millikan assumes to be that of guiding bacteria safely through water; and given how C responds to the pull of a magnetosome state, C will perform its Proper Function (in the Normal way), only if the pull is in the direction of oxygen-free water. Like the thermostat states, the purpose—and hence the content, if such there be—of magnetosome states is determined by the purpose(s) of their consumers. Thus, current pulls not in the direction of oxygen-free water are mistakes, according to Millikan, even if such pulls are in the direction of magnetic north. But despite its attractions, Millikan's account cannot be right.

III. Teleology and Discrimination

Consider the following fictional example:

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

There are several candidates for the content of a (current) kimu's B-token: Lo, redness; Lo, Wavelength W; Something nice is over there; There's that nice mountaintop again; etc. Or perhaps there is no determinate content of B-tokens to speak of. But one thing strikes me as obvious: B-tokens are not about snorfs. Nonetheless, this is the consequence that Millikan's account delivers. Indeed, in reply to this example—which I offered elsewhere to serve a different purpose—Millikan insists that, for a kimu, a B-token "signifies roughly, 'fewer

snorfs this way.'"[11] I address the indexical aspect (*this* way) of B-tokens presently; but the idea, according to Millikan, is that a B-token represents a certain spatial location as being relatively snorf-free, when compared with the kimu's present location. Let me first make clear why Millikan's account has this consequence. Then I'll say why the consequence is unacceptable.

The kimu example is designed to be just like that of the frog, with one exception: Whereas bugs at least sometimes cause neural firings in frogs, snorfs never cause B-tokens in kimus. But instantiations of a property don't have to causally affect a mechanism to be selectively important with respect to the proliferation of the mechanism. There is a famous species of moth in industrial England, whose members used to match the color (white, more-or-less) of trees in their niche. The trees have become soot-covered; but the moths have also changed, so that current moths match the color (dark grey, more-or-less) of current trees. Tree color was obviously important with respect to the selection for the current genetic mechanism controlling moth-color. But tree color has never caused any moth, former or current, to have the color it does (cf. chameleons). The same example also makes it clear that organisms need not be able to distinguish things that have property P from things that don't in order for instantiations of P to be selectively important to the organisms. The fitter (grey) moths didn't, in any ordinary sense of the word, distinguish soot-covered from white trees; they just matched the soot-covered trees.

Similarly, instantiations of "snorfness" can and do explain the proliferation of the kimus' mechanism M without B-tokens ever being caused by snorfs, or any kimu being able to distinguish snorfs from non-snorfs. All the selection explanation requires is that B-tokens sometimes be produced when a kimu's immediate locale is becoming snorf-infested. Millikan's conditions for having a Proper Function are satisfied. For (1) some ancestors of M produced B-tokens when a snorf-free zone was "over there," and production of these B-tokens helped former kimus get to a snorf-free zone; (2) there was some property Q of these M-ancestors (viz., their sensitivity to a certain wavelength of light), such that having Q sometimes caused said M-ancestors to produce B-tokens when a snorf-free zone lay "over there," with the result being a positive correlation, over some group of individuals including (but not limited to) said M-ancestors, between having Q and production of B-tokens when a snorf-free zone lay "over there"; and (3) citing (1) and (2) figures in a correct explanation of why current kimus have mechanism M, and hence, why M exists in kimus.

Putting Millikan's complex definition aside, former kimus with M did better at surviving and reproducing than kimus without M *because* those with M sometimes formed B-tokens when there were fewer snorfs "this way;" where "this way" is specified (indexically) by the particular

B-token produced, and contrasted with the general region occupied by the kimu at the time when the B-token was produced. This indexical element of a B-token's content is to be expected, given that the *adapted* Proper Functions of the mechanism that produces B-tokens will change, given *different* surroundings—just as the adapted Proper Function of a xerox machine will change, given different inputs. So, for a current kimu, M has the Proper Function of producing B-tokens when fewer snorfs lay "this way." We can also grant, with Millikan, that the Proper Function of M (and hence, of B-tokens) depends on the Needs of B-token consumers. For assuming that the relevant consumers of B-tokens in kimus are the mechanisms that guide kimu-behavior, and assuming (as Millikan does in the magnetosome case) that such mechanisms have the safe conduct of kimus as their Proper Function, then what B-token consumers Need when they consume B-tokens is that the region indexically specified by the particular B-token be relatively snorf-less.

It is important to bear in mind that the distinction between causation/discrimination on the one hand, and historical importance on the other is precisely what supports the rather nifty accounts of error provided by any teleological account of content. For recall that the general form of such explanation is the following: Some mechanism M produces B-tokens when and only when property Q is instantiated; but M is supposed to produce B-tokens when and only when property P is instantiated; so false beliefs are produced when Q but not P is instantiated. In this schema, Q depends on what can cause B-tokens; or again, it depends on what property the system can discriminate in virtue of having M. P depends on the purpose of the mechanism as established by historical circumstances. So the possibility of cases in which B-tokens are supposed to indicate something other than what causes them is not an unimportant spandrel of Millikan's view. All cases of false belief are held to be cases of this type. The kimu case simply brings out the following corresponding feature of Millikan's account: It may be that *true* beliefs are supposed to indicate something other than that which actually causes them as well.

Indeed, Millikan relies on this last point in her analysis of the magnetosome case. For oxygen-free water never causes the magnetosome to be in any state whatsoever. Because of a *correlation* between magnetic north and oxygen-free water in the bacteria's local environment, a mechanism that can produce states that reliably (if imperfectly) covary with magnetic north is a useful thing for bacteria.[12] Similarly, because of a local correlation between the presence of red things and the lack of snorfs, a mechanism that produces states that reliably (if imperfectly) covary with the presence of red things is a useful thing for kimus. But just as the connection between magnets and magnetosomes isn't what explains the proliferation of the latter, the connection between red things and the

kimu's mechanism M isn't what explains the proliferation of the latter. Moreover, if one insists that the connection between red things and the kimu's mechanism M does explain the proliferation of the latter, then one owes an account that avoids the inference, *mutatis mutandis*, to the parallel claims that: appeal to magnets explains the proliferation of magnetosomes; appeal to black specks (where both bugs and pellets count as specks) explains the proliferation of the frog's neural mechanism; etc. One cannot simply grant these parallel claims. For that would be to undercut the very account of error that makes teleological accounts attractive.

Millikan thus bites the bullet, holding that kimus' B-tokens are about snorfs. But then kimus climb the hill *because they believe* that the hill is snorf-less (and, presumably, they also want to avoid snorfs). Moreover, when kimus move toward red things on other occasions (on the flat), they are *acting on the belief* that the area in question is snorf-less. But such intentional explanations of kimu-behavior are about as implausible as intentional explanations can be. We have no reason, other than Millikan's theory, to think kimus have any beliefs or desires about snorfs. Indeed, we have overwhelming reason to think the contrary. Put a snorf in front of a kimu, and the kimu will not—unless a certain wavelength of light is also present—run away; and if snorfs happen to be red, then according to Millikan, kimus will think that packs of snorfs are snorf-free zones. Kimus can, at least under many circumstances, reliably discriminate red things from blue things; but they can't reliably discriminate snorfs from non-snorfs. Kimus, we may suppose, quite literally wouldn't know a snorf from a hole in the wall. We could collect a great deal of evidence in support of these claims by running standard psychological tests, i.e., by presenting kimus with various pairs of stimuli and checking for statistically significant patterns in their behavior. Simply put, all the behavioral evidence is going to tell against the claim that kimus have beliefs about snorfs; and this is surely relevant to the question of what beliefs kimus have.

My claim here is not a verificationist one. For I don't say it is impossible that kimus have snorf-beliefs given the considerations above. I just say that Millikan's account yields very implausible intentional explanations of kimu-behavior; and this result is significant. For what are intentional states, if not those states that figure in (correct) intentional explanations? Of course, implausible explanations *may* be correct; but plausibility is all we have to go on. Millikan's emphasis on historical importance rather than discriminatory ability would also lead to missing apparently good intentional explanations. Suppose a system has an innate disposition, in virtue of which it reliably produces s-tokens when conditions C, and s'-tokens when conditions C' obtain; but suppose that there is no history of selective advantage in having the

capacity to distinguish C from C'. Millikan's theory cannot explain any semantic difference between s-tokens and s'-tokens. But humans surely have discriminatory capacities that have never been selectively advantageous, in terms of either biological evolution or reinforced learning. Consider perfect pitch. If Jerry's capacity to discriminate a well-tuned A from a slightly flat one has no selective history, Millikan's theory will rob us of intentional explanations like: Jerry winced because he noticed that the orchestra was flat. One can deny that there is any correct intentional explanation to be had here. But the pile of bullets is growing, and there are still more.

Suppose that no kimu has ever seen (heard, smelled, etc.) a snorf, because if a kimu gets close enough, it gets eaten before a single snorf-caused neuron fires. One needn't hold that beliefs are about what causes them to hold that, in the absence of *any* causal interaction between B-tokens and snorfs, B-tokens are not—indeed, cannot be—about snorfs. It will also turn out a kimu never forms a B-token at time t *because* there yonder lies a snorf-free zone at t. That is, it is never snorfs that make kimus think about and avoid snorfs. In this sense, kimus form true B-tokens only by accident on Millikan's view. Now it is clearly possible for some tokens of a belief type to be true accidentally. But I find it difficult to imagine that *all* true tokens of a belief type can be true accidentally. Perhaps my intuitions here are products of causal theories of reference and reliabilist theories of knowledge. But this is not to say the intuitions are wrong; nor is it to say that the cost of giving them up is low.

One can, of course, claim that a theory of content will be somewhat revisionary of our common-sense practice of providing intentional explanations; and I have no objection to theory changing practice. But explaining kimu-behavior by saying that kimus believe there are fewer snorfs up the hill is radically revisionary; and Millikan hasn't offered any independent motivation for such radical revision. Moreover, the revisionist acquires the burden of providing a motivated set of constraints on a theory of content. That is, one has to say what counts as providing a correct theory of *content* (as opposed to a characterization of some other relation R that mental states might bear to the environment); and this becomes very hard—perhaps impossible—once one rejects the constraint that a theory of content has to yield plausible intentional explanations of behavior. Or to put the point another way, if the kimu-example doesn't count as evidence against Millikan's theory, then I want to know what would so count.

One might admit that the kimu-example—or better, the set of examples it represents—counts as *prima facie* evidence against Millikan's view, while maintaining that her theory is the best available theory of content, and so ought to be accepted on those grounds. I'm sympathetic to this

line of reasoning in general; although assessing competing theories is not possible here. But one might challenge the assumption that one needs a reductionistic account of the kind Millikan provides; and if Millikan's is the most plausible account of its kind, then perhaps no such reduction-istic account is correct. But I cannot press this line of argument here, and will content myself with the claim that the kimu-example provides evidence—strong evidence, I think—against Millikan's theory. It is also worth noting that the problem raised by the kimu-case is not peculiar to Millikan's teleological theory. For if natural selection explanations can ever support the "because-claims" required by TCC, then kimus have their B-token producing mechanisms *because* former tokens of the mechanism-type sometimes produced B-tokens when there were fewer snorfs "over there"; and so we have reason to reject TCC. But if selection explanations cannot support the "because-claims" of TCC, then we have no reason to think that teleological accounts of content apply to non-artifacts.[13]

IV.   Selection and Intention

The point of the last section can be summarized quickly: In focusing on the *biological* function of intentional states, Millikan has lost track of the *theoretical* function of intentional states, viz.. providing plausible expla-nations of the behavior of intentional systems. Nonetheless, Millikan has a defined *a* relation, call it 'PF', that beliefs can bear to properties that are sometimes instantiated in the system's environment. We have good reason to think that PF is not the relation "has as its content that." But sometimes, perhaps often, a belief bears both of these relations to the same property. Or put another way, selection explanations and inten-tional explanations sometimes make reference to the same properties; and this is probably what makes teleological accounts of content like Millikan's look so plausible. But even when, as in the frog's case, both kinds of explanations appeal to bugs, one is providing quite different kinds of explanations of the relevant frog-behavior, i.e., tongue-snapping. For we need to distinguish (i) ethological explanations that show how a certain behavior plays a useful role in the life of a system (or more generally, the members of a species), and (ii) intentional expla-nations that show how a certain behavior is the product of how the system *takes the world to be*. A non-biological example will be useful here.

   Suppose that, at time t, the articles and editorials published in a leading newspaper all have a distinctively conservative slant. If one asks why the paper has the political slant it does, one might get (at least) two different kinds of answers. Call the first the "powers-that-be" expla-

nation: The publishers and/or advertisers and/or other powers-that-be have conservative political views; as a result, journalists and editors who write pieces without a conservative slant get fired; and the current staff is composed of those who "survived," *viz.*, those who write pieces with a conservative slant. Call the second the "ideological-staff" explanation: The current staff is composed of political conservatives, and they express their political views in their writing. Both explanations can be correct explanations of why the newspaper has the slant it does. For it may be that, not only do staff members write with a conservative slant, they do so intentionally and sincerely; and the "powers-that-be" explanation might well explain why the current staff is composed of ideologues. But the "powers-that-be" explanation does not assume the sincerity of the staff. Liberal staff members may conceal their personal views. Indeed, the "powers-that-be" explanation doesn't even assume that the various journalists and editors *intentionally* write right-leaning articles. Some staff members may unwittingly "fit the bill" without ever thinking about it, consciously or otherwise.

   We know, then, that the "powers-that-be" and "ideological-staff" explanations are different. For the former can be true while the latter is false. Similarly, the latter can be true while the former is false. The ideologues may have been the only applicants for staff positions, much to the dismay of the liberal publisher. Moreover, the "powers-that-be" explanation is historical and statistical; while the "ideological staff" explanation is ahistorical and causal. The former explanation no doubt requires appeal to intentionality at many points: the intentions of the "powers-that-be"; the general intentional states required to write an article of any kind; etc. But the "powers-that-be" explanation does not require appeal to the claim that the journalists and editors *take the political world* to be a certain way. The "ideological-staff" explanation differs in exactly this respect. It depends crucially on (i) the journalists and editors having conservative views, and (ii) the conservative slant of the paper being the result of (i). Thus, we have two quite distinct expla-nations of "why the paper is the way that it is." To make the Davidsonian point: Neither explanation can be substituted for the other; for to move from one kind of explanation to the other is to change the subject.

   Similarly, I suggest, one can provide an ethological explanation of a given frog's tongue-snapping behavior by citing the history of selective advantage behind tongue-snapping *vis-à-vis* bugs; and one can provide an ethological explanation of a given kimu's hill-climbing behavior by citing the history of selective advantage behind hill-climbing *vis-à-vis* snorfs. That is, one can provide *an* explanation of this behavior by noting (together with, *inter alia*, facts about genetic control and herita-bility) that such behavior has historically been useful for members of the species. One may be able to provide a plausible intentional explanation

of the same frog-behavior by citing bugs as well; and if so, well and good. But it doesn't follow that *in general* the properties cited by the ethological explanation will figure in plausible intentional explanations. The kimu case provides a counterexample to this claim. Similarly, the newspaper example shows that ethological and intentional explanations are different kinds of explanation. Explaining why a system has *historically come to behave* the way it does is one thing; explaining why the system *now behaves* the way it does is another thing. The latter is the domain of intentional explanation; and to move from intentional to ethological explanation is to change the subject. If ethological explanations involving mental states can be given, no doubt they will advert to some relationship between mental states and the environment; and perhaps Millikan's definition of 'Proper Function' captures this relation. If so, then Millikan has sketched a new and potentially interesting kind of explanation of a certain range of behavior.[14] But this is not intentional explanation; and the relation captured by Millikan's apparatus is not the relation that intentional states bear to their contents.

One last point on this issue. It is striking that ethological and intentional explanations of behavior often advert to the same properties in the system's environment; and one wants an explanation of this fact. Millikan explains the coincidence of explanations by holding that intentional explanation is a species of ethological explanation. But there is another explanation readily available. The ability to represent the world is of tremendous biological value. So one would expect systems that represent the world to flourish. One needn't say that mental states are selectively advantageous in virtue of their intentional properties, as opposed to their effects on behavior; for intentional states do affect behavior. Neither need one say that intentional properties, as opposed to discriminatory abilities, figure in selection explanations. All one needs to say is that systems with the discriminatory capacities relevant to having intentional states (whatever these discriminatory capacities are) are more likely to be fitter than systems lacking such capacities. Thus, there is likely to be selection in favor of systems that have the ability to form intentional states. So one *expects* that, often anyway, there will be ethological explanations of behavior that is also explained by appeal to intentional states. But it is not that ethological explanations are, in any sense, prior (or identical) to intentional explanations. On the contrary, since natural selection occurs over generations, we should expect that intentional explanations often precede ethological explanations; and at a minimum, intentional explanations are independent of ethological explanations.

One might have been inclined to reject Millikan's theory at the outset on the following grounds: We can imagine holding a system's current behavioral dispositions and discriminatory capacities fixed, but varying its evolutionary history so that natural selection played no role in the

proliferation of these dispositions and capacities; in such a situation, the system would have all the same intentional states. Such arguments are, of course, not conclusive. For one can (*à la Kripke*) try to explain away such intuitions as a reflection of epistemic possibility, while holding that it is metaphysically impossible to vary the evolutionary history in the way imagined without stripping the state of any content whatsoever. But arguments based on epistemic possibility sometimes have true conclusions; and this is one of those times.

McGill University
Montreal, Quebec
Canada

NOTES

[1] For helpful comments and discussion, thanks to: Ned Block, David Brink, David Davies, Sue Dwyer, Eric Lormand, Georges Rey, and Bob Stalnaker.

[2] Error has been recognized to be a puzzler at least since the *Theatetus*. Recent discussions of rule-following have sparked interest in accounting for the sense(s) in which meaning is normative; see Saul Kripke, *Wittgenstein on Rules and Private Language.* (Cambridge: Harvard, 1982.) Paul Boghossian, "The Rule-Following Considerations," *Mind* 98, pp. 507–49 (1989) discusses the implications of these issues for theories of mental content.

[3] Ruth Millikan, *Language Thought and Other Biological Categories.* (Cambridge: MIT, 1984); "Thoughts Without Laws; Cognitive Science With Content," *Philosophical Review* 95, pp. 47–80 (1986); "Biosemantics," *Journal of Philosophy* 86, pp. 281–97 (1989); "Clarifications on *Language, Thought, and Other Biological Categories,*" *Annals of Scholarship* 7, pp. 147–49 (1990); and "Explanation in Biopsychology?" John Heil and Alfred Mele (eds.), *Mental Causation* (Oxford, 1993). For other teleological accounts of content, see: David Papineau, "Representation and Explanation," *Philosophy of Science* 51, pp. 550–72 (1984); Daniel Dennett, *The Intentional Stance.* (Cambridge: MIT, 1987); Fred Dretske, *Explaining Behavior.* (Cambridge: MIT, 1988); Mohan Matthens, "Biological Functions and Perceptual Content," *Journal of Philosophy* 85, pp. 5–27 (1988).

[4] One might want an adequate account to reveal what tokenings of 'F' and 'E' have in common, viz., the purpose of indicating the level of fuel in the tank. As we shall see, Millikan's theory does just this. A side point: My car can have its fuel gauge as a part because (1) my fuel gauge sometimes registers 'F' when the tank is full, (2) other tokens of the fuel-gauge type sometimes register 'F' when the tank is full, or both, depending on whether or not my fuel gauge is "quality tested" prior to release from the factory.

[5] In the frog's case, Q is instantiated (at least) sometimes when bugs are present, sometimes when dark metal pellets are present, and other times as well—e.g., when an evil demon/experimenter shines a light on the frog's retina. To make the familiar point: Property Q can be happily expressed in English only as something like "the property such that the frog tokens its belief when and only when the property is instantiated"; other attempts to characterize Q are likely to require (very) long disjunctive predicates.

[6] In general, any such account presupposes that we can always make sense of different creatures having tokens of the *same mechanism type,* without begging the question. (But let us grant this.)

[7] I omit Millikan's technical definition of 'ancestor', taking the notion to be intuitively clear. The definitions for all of Millikan's central theoretical terms are given in her book (1984), *op. cit.*, pp. 19–43.

[8] For further discussion, see Millikan (1989), *op. cit.*

[9] Fred Dretske, *Knowledge and the Flow of Information*. (Cambridge: MIT, 1981.)

[10] Millikan (1989), *op. cit.*, p. 291.

[11] Millikan (1990), *op. cit.*, p. 149. I originally used the example in Paul Pietroski, "Meaning and Evolution," *Annals of Scholarship* 6, pp. 318–330 (1990) to suggest a certain "interest-relativity" of selection explanations that might infect Millikan's account of content. Putting aside this point, which I don't wish to pursue here, Millikan's account is unequivocal (as she insists): the kimus' B-tokens mean 'fewer snorfs this way'.

[12] Appeal to the "local" environment is crucial, as there are different *kinds* of correlation between oxygen-free water and magnetic north. Magnetosomes in Southern hemisphere bacteria repel the bacteria from the surface, as opposed to pulling them toward the bottom.

[13] One might develop a "hybrid theory" that takes both natural history and causal history into account: Dretske (1988), *op. cit.*, may be an example. I cannot address such accounts here. But I suspect that any theory capable of handling kimu-like cases will not require appeal to teleology at all; and in any case, hybrid theories would be very different from Millikan's theory.

[14] See Millikan (1993), *op. cit.*, for a picture of what "Biopsychology" might look like. There is a vexed question here: How much can a theory T be "corrected" and still count as *the same theory* for purposes of reducing T to another theory T'? There will, of course, be unclear cases; but for reasons given in the text, I think the "corrected" version of intentional psychology that Millikan requires amounts to a new theory. Or again, if the only options vis-à-vis the intentional are reduction and elimination, then I think Millikan's route amounts to eliminating the intentional and *replacing* it with the ethological. Of course, it's far from obvious that reduction and elimination exhaust the alternatives here.

THE FOUNDATIONS OF THE CALCULUS AND THE CONCEPTUAL ANALYSIS OF MOTION: THE CASE OF THE EARLY LEIBNIZ (1670–1676)

BY

MICHAEL J. WHITE

I. Introduction

My purpose in this paper is to examine critically two of Leibniz's early conceptual analyses of continua, in general, and of motion, in particular, in connection with his mathematical maturation during the early 1670's. The earlier of these two analyses occurs in the *Theoria motus abstracti (TMA)*, which was written in the winter of 1670–71 when Leibniz—according to his own later account—was mathematically self-taught and callow.[1] The later, the dialogue *Pacidius Philalethi (PP)*, was written in October of 1676. In the meantime, Leibniz had received advice concerning his mathematical reading from the English mathematician John Pell, had entered into correspondence concerning mathematical matters with John Collins and Henry Oldenburg of the Royal Society, had sojourned in Paris where he came under the mathematical tutelage of Huygens, and had purchased and probably read at least parts of the lectures by Isaac Barrow on optics and on geometry. It is also during the latter part (1674–76) of this period that Leibniz made major contributions to the development of the differential and integral calculus.

Despite Leibniz's immense mathematical maturation during this period, he was in 1676 still wrestling with fundamental conceptual or ontological problems concerning continua, in general, and motion, in particular,