

# A Defense of Millikan's Teleosemantics<sup>\*</sup>

Timothy Lane, Associate Professor  
Department of English  
National Chengchi University

(收稿日期：1998年8月7日；修訂日期：1999年4月30日；授受刊登日期：1999年9月27日)

## Abstract

Typical of the criticisms leveled at Millikan's teleosemantics are Pietroski's charges: (1) under Millikan's theory it can be the case that items of content are completely disconnected from that which causes them; (2) Millikan fails to identify conditions that would count against her theory; (3) her view is radically revisionary but she provides no independent motivation for it; (4) she ignores the role unselected discriminating abilities may play in intentional explanation; and, (5) she conflates teleological and intentional explanation.

Although Pietroski's criticisms, especially (1), have a measure of intuitive appeal, I argue that they are less substantial than they might appear and that they do not constitute a serious challenge to Millikan's semantic theory. I defend Millikan's position through conceptual analysis, thought experiment, and data from evolutionary biology. To demonstrate how teleosemantics can be expanded, I attempt to show it is applicable not just to cases of evolutionary selection, but also to human learning. Furthermore, I identify places where revision of Millikan's theory might be in order and places where we are unlikely to make substantial progress until we achieve certain conceptual refinements and a greater understanding of relevant empirical phenomena.

Key word : semantics, teleology, mental representation, evolutionary selection, explanation

---

<sup>\*</sup> I thank Josefa Toribio and three anonymous reviewers for their helpful comments on previous versions of this manuscript.

## INTRODUCTION

Teleosemantics attempts to give a teleofunctional account of what determines the semantic content of mental representation. Although there is likely no way to explicate the principle concepts in a manner that is both brief and non-contentious, roughly it can be said (at least for the works of Dretske, Millikan, Neander, and Papineau) that advocates are attempting to develop an account of mental representations that treats them as evolved, biological complexes, along the lines of, say, Darwin's treatment of variation among the finches of the Galapagos islands. Generally speaking (for expository purposes I here pass over a few fine distinctions), they adhere to the prevalent view that the semantic content of a mental representation is what is picked out by "that" clauses in constructions of the type: "Mr. Chen believes that Taiwan is autonomous," or "Mr. Ma hopes that his hair will not fall out." Such indicative sentences with "that" constructions are typically referred to as propositional attitudes. Beliefs and other attitudes (e.g. hope, desire, expect, and fear) can be "about" things—e.g. Taiwan autonomy or mayoral baldness. To say of a propositional attitude that it has the property "aboutness" is to say that it is meaningful, that it has semantic content. And, the property of propositional attitudes whereby they are "about" things (likewise, whereby words or sentences might be "about" things) is typically referred to by philosophers as intentionality.

Teleosemantics, especially the teleosemantics as has been developed by Ruth Millikan for over a decade, aspires (1) to naturalize semantics, (2) to give semantics an ecological orientation, and (3) to provide solutions for problems that continue to vex rival semantic theories. (1) The desire to naturalize is motivated in part by the belief that "aboutness" is not a fundamental property of the universe: it is neither primitive nor unanalyzable. Naturalists strive to give an empirically adequate and explanatory account of how "aboutness" can emerge from non-semantic properties, properties that are more fundamental.

(2) The desire to provide an ecological orientation is motivated by the belief that the mechanisms humans use to form propositional attitudes are biological. Accordingly, they have biological functions. Since biological functions can only be determined with reference to evolutionary history in a normal environment (e.g. tree rings “normally” correlate with tree age, unless climatic conditions are “abnormal,” or unless disease—an “abnormal” condition--has infected the tree), we stand to gain little from confining investigations to analysis of input-output series under laboratory conditions.

And, (3) we already have some indication of teleosemantic success in resolving a problem that has plagued other semantic theories, the problem of misrepresentation (or, sometimes called the disjunction problem). Briefly, consider that mental representations in us that co-vary with the presence of hair will also co-vary with the presence of hair *or* very natural-looking toupees. The problem for semanticists (at least those of a certain theoretical disposition) then becomes explaining how it can be that mental representations can have the content “hair” and not “hair *or* toupee.” A teleosemanticist holds that hair-beliefs stand for “hair” and not “hair *or* toupee *or* weave” because it is their purpose to be held when hair, but not toupees (or weaves, or whatever else might look like hair) are present.

But teleosemantics is not without its problems. Typical of these problems are those raised by Paul Pietroski (1992) in an influential critique of Ruth Millikan’s version of teleosemantics. Because Millikan has developed the most detailed and systematic version of teleosemantics, were Pietroski’s criticisms to go unanswered, it might prove debilitating for future theory development and an obstacle to promoting theory influence among semanticists of alternative theoretical orientations.

Therefore, I here attempt a limited, provisional defense of Millikan’s teleosemantics—a defense against the criticisms raised by Pietroski. I do not claim to be able to decisively refute Pietroski, but I do hope to have succeeded in sketching responses that an advocate of

teleosemantics can use to counter and diminish the force of his criticisms, criticisms which have an admittedly *prima facie* plausibility.

Specifically, Pietroski (1) is troubled that under Millikan's theory it can be the case that items of mental content are wholly disconnected from that which causes them--Millikan even advocates "ignoring" the causes of representations. In an attempt to refute this idea, an idea which Pietroski rightly regards as critical to Millikan's project, he develops a thought experiment which is intended to show that on Millikan's view mental content is just "accidental." Because his thought experiment has an undeniable intuitive appeal for those of competing theoretical positions and because it has been heralded by some critics (e.g. Rey 1997: 248-249) as "perhaps the most serious worry" concerning teleosemantics, the main body of my paper, section three, is devoted to an attempt at mitigating this worry and thereby diminishing its force.

Pietroski also observes that (2) Millikan has not identified conditions that would count against her theory; (3) Millikan's view is radically revisionary yet she provides no independent motivation for it; (4) she ignores the role unselected discriminating abilities may play in intentional explanation; and (5) she ignores the difference between historical (how a system has come to behave the way it does) and ahistorical (why a system now behaves the way it does) explanations--in so doing she appears to conflate teleological and intentional explanation. I also respond to these criticisms and try to vitiate the worry that attends to them. But along the way I do indicate points at which some minor revision of Millikan's views might be in order, and points at which we will likely need to await the results of further conceptual clarification and empirical research. My proposed revisions are not likely to succeed in converting Pietroski, but I believe they diminish Millikan's vulnerability to criticisms of the type devised by Pietroski.

In section one, I briefly present the relevant features of Millikan's views and draw attention to a characteristic of her teleosemantics that distinguishes her from other teleosemanticists, most notably Dretske. In section two, I introduce Pietroski's worries (1) through (4) and give an

abbreviated version of the thought experiment that he takes to count strongly against Millikan. In section three, I defend Millikan against Pietroski's charge that her view makes mental content appear to be "accidental": (a) I argue that the only sense of "accident" to which Pietroski can appeal is problematic and that it has limited force in this context. (b) I also show that, by contrast, there is a clear sense in which, under Millikan's view, content is non-"accidental." And, (c) by developing three thought experiments I reveal an ambiguity in Pietroski's thought experiment, an ambiguity that once realized helps resist the appeal of his conclusion and helps with the discovery of an alternative conclusion. Moreover, I argue that there are clear circumstances which would show Millikan's view to be inadequate, were they to obtain; and, I make a first pass at providing independent motivation for her view. In section four, I indicate my worries about thought experiments of this type, especially as applied to this problem, and both to relieve those worries and to try to expand the range of application for Millikan's view, I analyze an event from the history of science, an event that I argue is analogous to Pietroski's in all relevant respects. I contend that, at least where natural kinds are concerned, my example shows that Millikan's theory can apply to human learning. In this section I also further develop reasons that may provide independent motivation for Millikan's view and I suggest that Pietroski's unselected discriminating abilities may be essential to some explanations, but that it is unclear whether they are needed for intentional explanations.

In section five, I present Pietroski's (5) worry that Millikan conflates historical, teleological explanation with ahistorical, intentional explanation. Pietroski distinguishes between the two and argues that intentional explanation is "independent" of teleological explanation. Section six tries to cast doubt on the viability of this distinction, as well as to suggest that such a distinction might prove an obstacle to studies of human learning.

## SECTION I

In Millikan's (e.g. 1989, 1993, 1994, 1995, and 1999) teleosemantics the categories of intentional psychology are treated as categories with biological purposes/functions.<sup>1</sup> She distinguishes clearly between what something is designed to do, its proper function (which can come about either through the intentions of a designer or through a mindless process like natural selection), and how it does whatever it does. Hair-beliefs represent hair, not toupees, because it is their purpose to do so; they have been designed to be about hair. Just how they accomplish this is a separate matter. Moreover, she emphasizes that it is not essential that a thing always, or even typically, perform its proper function (e.g. few sperm ever actually fertilize an ovum).

It is a distinctive characteristic of Millikan's view that she focuses on the "consumption" rather than the "production" of mental representations. That is<sup>2</sup> it "is the devices that use representations which determine these to be representations and ... determine their content. If it really is the function of an inner representation to indicate its represented... (it must function as a representation) for the system itself." For the naturalist who aspires to describe intentionality, representation production is immaterial (Millikan 1994: 247). All representations are produced by systems whose proper function is to make the representations "correspond by rule to the world." But the rule of correspondence is determined entirely by the consumers.

Millikan clearly distinguishes her views from the views of those who advocate causal/informational accounts of semantics. By way of example, she (1994: 251-252) cites Dretske's (1994) discussion of the magnetosome, the magnetic-field-sensitive organ which is found in certain northern hemisphere bacteria. Because of the magnetosome these bacteria are

---

<sup>1</sup> Millikan (e.g. 1989: 293-294) explicitly links the notion of purpose to function.

<sup>2</sup> See Millikan (1994: 245). For a similar view of teleosemantics see Papineau (1990 and 1993: 55-101). Dennett's (e.g. 1993: 222-227) views are also very similar, but Dennett (p. 224) finds that his "intentional stance" must even be taken toward the evolutionary process. Millikan disagrees. Moreover, though both Millikan and Dennett are

pulled toward geomagnetic north--i.e. down, away from surface water, away from oxygen, which is toxic to these bacteria. Although the direction of oxygen-free water is not a factor in causing the direction of pull on the magnetosome, on Millikan's account the function or purpose of the magnetosome is to bring it about that bacteria move into oxygen-free water. Accordingly, the pull of the magnetosome represents the location of oxygen-free water.<sup>3</sup> "What the magnetosome represents is only what its consumers require that it correspond to in order to perform their tasks. Ignore... how the representation... is normally produced. Concentrate ... on how the systems that react to the representation work..." (Millikan 1994:250).

For Millikan (1994: 256)<sup>4</sup> "natural selection is not the only source of proper functions." Design can be achieved both through biological evolution and through learning or experience (1994: 245 and 251). Moreover, just because something has been designed to achieve a purpose does not mean that its presence will normally or even often result in achievement of that purpose. Selection does not require high frequency, only sufficient advantage (recall the sperm-ovum example). Beaver tailsplashing, in "the beaver splash semantic system," means danger. But it may be the case that most splashing does not occur in circumstances of actual danger.

## SECTION II

Pietroski (1992: 273-278) argues that Millikan's move of trying to develop a naturalistic account of semantics by ignoring that which "produces" or causes the inner representations is *prima facie* implausible. To help make his point, he develops a thought experiment (p. 273):

---

intentional realists, their senses of realism are very different.

<sup>3</sup> Compare Dretske (1994: 167): "It sounds a bit far-fetched ... to describe the bacteria's sensory mechanism as indicating, and having the function of indicating, the whereabouts of oxygen... This is, after all, a magnetotactic, not a chemotactic, sensor." Dretske's sense that Millikan's account is "far-fetched" exemplifies the intuition to which Pietroski appeals through his thought experiment.

<sup>4</sup> Papineau (1990: 37 and 1993: 59) makes the same point. Also see Millikan (1995: 214 and 218).

The kimus live near a ... hill. Their only predators are snorfs... who roam past the hill each morning. Kimus used to be "color blind." But in virtue of a genetic mutation, one ... kimu ... Jack... came to have an internal mechanism *M* that produced tokens of a physically specifiable 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 ... Jack (like his descendants) turned out to have a "fondness" for things red; i.e. other things being equal, Jack would move towards the distal causes of *B*-tokens when such tokens were produced... each morning, Jack trudged up the hill and ... avoided the snorfs. Natural selection took over ... Jack's mechanism type proliferated throughout the species. There was no other reason... for the selection in favor of having the "color mechanism."

According to Pietroski (1992: 273-274), Millikan claims that for a kimu, a *B*-token "signifies roughly, 'fewer snorfs this way.'" Indeed, given her interpretation of Dretske's magnetosome case, to be consistent it seems she would have to regard the *B*-token as being about a place wherein there are fewer snorfs.

Pietroski (1992: 276) argues that since intentional states are states that figure in correct intentional explanations, and since Millikan's account yields implausible intentional explanations of kimu behavior, her account's implausible explanations should be taken as strong evidence that her theory is flawed. As examples of implausible intentional explanations, Pietroski cites the following:

Kimus climb the hill because they believe that the hill is snorf-less ... 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 ... if snorfs happen to be red ...



according to Millikan, kimus will think that packs of snorfs are snorf-free zones ... kimus ... wouldn't know a snorf from a hole in the wall.

And we could add to the list that were a magnet placed on the surface of the water above a magnetosome-carrying bacterium, the pull of the magnetosome would represent oxygen-free water--even though the water was oxygen-rich. Bacteria wouldn't know the difference between oxygen and lead.

Essentially, Pietroski is worried that it might never be snorfs that make kimus think about and avoid snorfs. It could be the case that all true B-tokens were true "accidentally." What's more Pietroski is bothered by a cluster of related worries: (1) Millikan's view appears to be an unfalsifiable, just-so story. Pietroski (p. 277) wants to know if his example wouldn't count against Millikan, exactly what would? (2) Pietroski (p. 277) also finds Millikan's view to be "radically revisionary" and he faults her for not providing "independent motivation." And, (3) he (pp. 276-277) feels Millikan's emphasis on history will lead her to miss good intentional explanations that advert to discriminatory abilities which appear not to have resulted from design.

### SECTION III

Pietroski invokes the concept of "accident," but offers no explication, even though this concept is clearly crucial to his criticism of Millikan's project. Without a clear, coherent sense of "accident," his main criticism rests on nothing more substantial than cotton candy or political promises. Apparently though he deems explication unnecessary, for the common sense notion of "accident" seems to evoke a strong intuitive response, so strong that not only are those unsympathetic to teleosemantics concerned (e.g. Rey 1997: 248-249), so too are some advocates (e.g. Dretske 1994: 167). But the common sense notion and the intuitive response that derives from it are importantly misleading.

I begin my reflections on Pietroski's concept of "accident" (henceforth, P-accident) by examining the three most common philosophical uses of "accident." The purpose of this exercise is to determine whether Pietroski does (albeit implicitly), or could (should the common sense view prove inadequate), avail himself of any one of them, in order to bolster his position. Alternatively, might it be the case that these philosophically, canonical uses of "accident" provide motivation for questioning the Pietroski usage? The first, because I name it after Hempel, I call an H-accident; the second, because I name it after Aristotle, an A-accident; and the third, because I name it after Gettier, a G-accident.

Since Pietroski is explicit in his concern that one matter which most bothers him is the implausibility of Millikan's intentional explanations, it is perhaps most appropriate to first consider how "accident" is canonically employed in philosophical accounts of explanation—those which I call H-accidents. The distinction between regularities due to physical law and regularities due to accidental concatenations of circumstances is often exemplified with cases of the following type: the regular falling of heavy bodies to earth seems not to be accidental for it seems explainable in terms of Newton's law of inverse squares and the three laws of motion. The equally regular phenomenon that all gold bars weigh less than 100,000 tons, however, does seem to be accidental.<sup>5</sup> It is often said that regularities of the first type can support counterfactual inferences and modal statements of physical necessity and impossibility, while those of the second cannot.<sup>6</sup>

But apparently Pietroski neither draws upon, nor could he draw upon the concept of H-accidents, for ex hypothesi the B-tokens function to keep kimus at a safe distance from predators. The regularities pertaining to kimus B-tokens can support some counterfactual inferences, and possibly even modal statements (though here we would require more

---

<sup>5</sup> For a thorough discussion of these examples and the problem of "accidental generalization" see Hempel (1991: 304-307).

<sup>6</sup> For criticism of this view see Cartwright (1983: 46-53).

background information) of necessity and physical impossibility. If anything, the concept H-accident leads us to be dubious of P-accidents.

So H-accidents, those most nearly associated with explanation, do not help Pietroski. What of A-accidents and G-accidents? The Aristotelian sense, A-accident, is intended to refer to the properties of things, without which, they could still exist (e.g. mayors can exist without hair, so hair is an accidental, not essential, property). How this would apply to Kimus is perhaps not obvious, but at least it can be safely claimed that discovery of a snorf-less zone is more nearly essential than it is accidental to their existence. So here too we find no support for Pietroski; on the contrary, we find additional grounds for doubt.

If we ignore Pietroski's stated concern with explanation and ignore worries about existence, it might be argued that G-accidents, at least on the surface, and in some versions, seem a bit more relevant to what Pietroski suggests with his thought experiment. A G-accident obtains when we believe things which are true, things which we can justify, but which nonetheless are not true in quite the way, or for quite the reasons, that we believe them. For example, standing in front of City Hall I form the belief that "a mayor of Taipei with a full head of hair is standing on the steps of City Hall." I form this belief because I can see Mr. Ma, a Taipei mayor with a full head of hair. I have no reason to doubt that that is real hair, for even in the wind and rain, or when tousled by female admirers, it responds like real hair. But, despite appearances, that is just a toupee. Still, there is "a mayor of Taipei with a full head of hair standing on the steps of City Hall," for just outside of my line of vision stands Mr. Chen, a previous mayor, who really does have a full head of hair. The upshot is that I have a belief that is true, and for which I can produce justification (of a sort), but many people would conclude that in this case I don't have knowledge. My justified belief is true, but true by accident, by G-accident.

G-accidents though do not lend support to Pietroski's view either, for they call into question the very relationship that Pietroski worries is missing from Millkan's account. Pietroski worries that kimus snorf-beliefs are not grounded in direct perceptual access to snorfs. But as

G-accidents show, direct perceptual access is not necessarily a guide to knowledge. If it does not reliably guide us toward knowledge, one may be reasonably inclined to wonder in just what sense it can be claimed to guide us toward semantic content.

Not only do H-, A-, and G-accidents fail to bolster Pietroski's position, they call it into question. Examination of H-accidents shows that, semantic oddity aside, the intentional explanations invoked in his thought experiment, do indeed support counterfactual inferences. Examination of A-accidents shows that there is indeed something essential (albeit not necessarily in a strictly Aristotelian sense) about that snorf-less area. And, examination of G-accidents raises doubts about the reliability of directly accessed perceptual knowledge. In short, philosophical accounts of "accident" do not lend support, to Pietroski's argument, they detract from it. At best we are left with nothing but an unexplicated common sense notion of direct, perceptual, causal access.

Having eliminated the possibility of drawing support from a more philosophically robust sense of "accident," let us proceed to consider the common sense view. Here I begin with a few rather pedestrian remarks about Pietroski's implied sense of direct, perceptual access. First, perceptions are not so direct as common sense may sometimes lead us to presume: for example, we do not see things directly. Rather, at the very minimum, patterns of light striking the retina play an intermediate role. Second, while traditionally it was assumed that perception occurs passively from sensory inputs, it is now generally accepted (e.g. Gregory 1994) that stored knowledge and assumptions actively affect most, if not all perceptions, even the most elementary. Third, in actual biological cases of the type described by Pietroski, there are likely to be additional internal variables. For example, squirrels (Calvin 1996: 22) preparing for the approach of winter are apparently reacting to longer nights. Even on an extremely simple model of this type, there is at least one internal, intermediary event: the dark precipitates secretion of melatonin by the pineal gland, and it is the melatonin, in turn, which triggers food hoarding. Four, there is some empirical evidence (e.g. Kornblith 1993: 61-82) which indicates that humans

have an innate propensity to conceive of the world in terms of unobservable essences; we apparently do not rely so heavily on observable features as common sense may suggest.

These remarks are not intended to imply that we are perceptually cut off from the world. Any animal, including *Homo Sapiens*, whose belief generating mechanisms do not afford sufficient cognitive contact with the external world, have in Quine's (1994: 66) words, "a pathetic but praiseworthy tendency to die before reproducing their kind." But with more than a touch of irony we can note that one of the best defenses for the common sense view of perception implied by the Pietroski thought experiment is an appeal to evolutionary selection, precisely the view that he finds so unsatisfactory as an account of mental representation.

Moreover, although the absence of direct causal connection is somewhat disconcerting, it should not be so disconcerting as to lead us to overlook what we do have. Pietroski describes a predator-prey relationship and their respective patterns of movement within a given niche. It may be troubling that this kimu belief about snorfs<sup>7</sup> is not fixed by any properties intrinsic to snorfs, but Pietroski has described a constant conjunction of entities (kimus and snorfs) and space-time arrangements (fixed time of day at which snorfs pass by and kimus move up the hill). Constant conjunction and cause have not yet been adequately disentangled in physics (e.g. Sklar 1993: 128-131), to press the issue here, at this early stage in the attempt to naturalize semantics, is, arguably, an unnecessarily high requirement. Further, all of the sciences acknowledge the explanatory value<sup>8</sup> of relational properties; it would be arbitrary to disallow them here. Just as the relation "taller than" might state an important relation between two organisms that have

---

<sup>7</sup> Even on Pietroski's limited account it remains an open question whether or not kimus may have other snorf-beliefs that might be fixed by intrinsic properties of snorfs. But then inadequate background is just one of the problems that make our judgements about such thought experiments unreliable.

<sup>8</sup> But relational properties are an occasional spark of controversy and they are more common and more welcome in some fields than in others. For a brief, elementary introduction to the distinction between intrinsic and relational properties, as well as some of the attendant implications, see Inwagen (1993: 33-37).

never crossed paths and who would fail to recognize, even react to, one another even if they did, so too might the relation “animal x goes where animal y is not.”<sup>9</sup>

Even should we choose to ignore these concerns, and allow that Pietroski is working with a defensible, intuitive sense of direct, perceptual cause, still it should not be forgotten that what strikes us as intuitively appropriate, on the basis of a priori reasoning, is not necessarily correct. Millikan treats “aboutness” not instrumentally, but realistically. She takes it to be a feature of the world, a feature concerning which our intuitions may well be crashingly incorrect—just as our intuitions concerning matter, space, and time have so often proved incorrect. If forced to choose between that which is empirically adequate and explanatorily powerful and that which most nearly coheres with our pre-theoretic intuitions, she would surely choose the former.

Millikan’s empirical orientation though does raise a legitimate concern with her view. Although Millikan frequently develops her ideas in the context of thought experiments not unlike those used by Pietroski, thought experiments which allow our intuitions concerning mental content to move freely from human to non-human species, this approach may rest on a mistake. Recently, Terrence Deacon (1997) has argued that the human mode of reference (what he calls “symbolic” reference) may be fundamentally different from the “nonsymbolic” (or “indexical”) mode of reference exhibited by other animals, including those to whom we are close in the phylogenic chain as well as those who have demonstrated intriguing linguistic abilities. One significant implication of his view is that there are not any simple languages, none that are extant (though our hominid predecessors presumably possessed them). He argues that we are descendants of a species that benefited from a “semiotic innovation” (Deacon 1997: 44), an innovation that was extremely difficult to acquire.

---

<sup>9</sup> I do not intend to downplay the significance of trying to disentangle cause from constant conjunction. But I am not overly sanguine about the chances for success. For one of the more popular efforts, explaining causal dependence in terms of counterfactual dependence, see Lewis 1993. I think Salmon (e.g. 1998) has made the most progress toward developing an adequate response to Hume’s challenge to find a physical connection between cause and effect; he takes as his points of departure the theory of relativity and a reorienting of attention away

If Deacon is correct, then to the extent that our judgements concerning mental content are motivated by intuitions that conflate the symbolic and the indexical (a distinction not unlike the philosophical distinction between sense and reference, Deacon 1997: 83) we may be misled by data on animal “languages,” and thought experiments that treat non-humans as capable of symbolic reference. To invoke a cliché, we may well be guilty of comparing apples and oranges; our reflections range continuously over phenomena that are not in themselves continuous. This worry provides part of the motivation for section IV, in which I try to apply Millikan’s view to a case of human learning.

Deacon’s empirical theory, however, is highly controversial and it rests on an idiosyncratic reading of the relevant data. For this and for other reasons I do not intend to use it to dismiss the Pietroski thought experiment—I think it sufficient that when we reflect on such thought experiments we bear in mind the possibility that “aboutness” is not necessarily of one kind, not necessarily a phenomenon that merely differs by degrees across the universe of animal species.

Even if we dismiss the Deacon worry, and put it aside, along with the other concerns cited above, still there is adequate reason to be dubious of Pietroski’s conclusions. These doubts can perhaps be most directly and effectively raised by working with Pietroski’s chosen tool of argument—the thought experiment.

Hempel (1965: 164-165) once usefully distinguished between those thought experiments (what he called “experiments-in-imagination”) which are more “theoretical” and those which are more “intuitive.” He treated the two as occupying end-points on a continuum. The former rest on explicitly stated principles and strict deductions (e.g. some of Galileo’s and Einstein’s are canonical); the latter rest on largely unarticulated principles, and inferences are made rather more casually. Since Pietroski does not ask that we violate any laws of nature (unlike many philosophical thought experiments) and he does describe a naturalistic setting, we needn’t place

---

or facts and toward processes, which he argues are of two types, causal and non-causal.

it at the wholly intuitive end of the continuum, but surely there are no grounds for placing it near Galileo's or Einstein's. Perhaps, being maximally fair to Pietroski, the best that can be said is that it occupies a gray, middle ground, tilting neither toward the theoretical nor toward the intuitive. Still, even thought experiments of this type can, as Hempel pointed out, assist with discovery. And I take it, on an appropriately charitable reading of Pietroski, that he claims his thought experiment assists with discovery of certain characteristics of mental content and attendant failings of teleosemantics.

Below I propose three thought experiments, similar to Pietroski's in content (all concern simple predator-prey relationships) and quality (i.e. all occupying roughly the same location between the theoretical and the intuitive). By means of some minor adjustments I hope to show that intuitions motivated by his experiment are not necessarily a consequence of resistance to the idea that mental representations are not caused by direct, causal, perceptual access. Resistance may well be motivated by other factors.

Thought Experiment-1 (TE-1): The kimus live near a hill. Their only predators are snorfs who roam past the hill each morning. Kimus used to be "color blind," but in virtue of a genetic mutation, one kimu, Jack, came to have an internal mechanism M that produced tokens of a physically specifiable type B in the presence of certain wavelengths of light. Snorfs, themselves dark gray, have a symbiotic relationship with cardinals whose color is just precisely that which causes Jack to form a B-token. Jack (like his descendants) turned out to have a "fear" of things red; i.e. other things being equal, Jack would flee the distal causes of B-tokens when such tokens were produced. Each morning, when the snorfs approached, with cardinals perched on their shoulders, Jack would flee. Natural selection took over. Jack's mechanism proliferated throughout the species. There was no other reason for the selection in favor of having the color mechanism.

Thought Experiment-2 (TE-2): The kimus had lived without predators for generations. Unbeknownst to them a colony of carnivorous snorfs had just immigrated to their island.



Kimus, having lived without predators for so long, are naturally inclined to be friendly. Kimus had been sensitive to only an extremely narrow range of sounds, and were wholly incapable of detecting high pitches (not even those of a human tenor). But Jack, a kimus, in virtue of a genetic mutation, came to have an internal mechanism *M* that produced tokens of a physically specifiable type *B* in the presence of certain high pitched sounds. Each morning, prior to setting out on their morning walk, snorfs turn, face a mountain, and sing a high-pitched yodel. The pitch of their voices, carried by echoes off the mountain produced *B*-tokens in Jack. Jack, like his descendants, turned out to have a “fear” of those high-pitched yodels; i.e. other things being equal, Jack would flee the distal causes of *B*-tokens when such tokens were produced. Lucky for Jack, because kimus found snorfs to be visually attractive and were naturally inclined to move toward them when they approached. The result: breakfast for the snorfs, carnage for the kimus. But Jack never saw the snorfs (and neither did his descendants), for since they are creatures of habit (e.g. they always yodel before taking their morning walk), Jack was never around to see them. Natural selection took over. Jack’s mechanism type proliferated throughout the species. There was no other reason for selection in favor of the “high-pitch mechanisms.” Also, it turns out that the snorf-yodel can be imitated by the songs of a flock of mockingbirds (mockingbirds have a talent for mimicry)—the two are indistinguishable to Jack—who occasionally fly past the kimus.

Thought Experiment-3 (TE-3): The kimus live near a hill. Their only predators are Egyptians who roam past the hill each morning. Kimus used to be “color blind.” But in virtue of a genetic mutation, one eccentric kimu, Moses, came to have an internal mechanism *M* that produced tokens of a physically specifiable type *B* in the presence of certain wavelengths of light. Each morning, something red on the hilltop—what to humans might look like a burning bush, but what was in fact just a granite outcropping reflecting morning sunlight--caused Moses to form a *B*-token when he looked up. Moses (like his descendants, who were mostly the offspring of a group of disaffected young) turned out to have a “fondness” for things red; i.e.

other things being equal, Moses would move towards the distal causes of B-tokens when such tokens were produced. Each morning, Moses trudged up the hill and avoided the Egyptians. Natural selection took over. Jack's mechanism proliferated throughout the species. There was no other reason for the selection in favor of having the "color mechanism."

I contend that TE-1 is "accidental" in the sense intended by Pietroski, with just one slight difference: here we can draw upon a theory of symbiosis to better understand the co-variation of cardinals and snorfs. Here, guided by the theory of symbiosis, the reader may find the claim that a B-token "signifies roughly, snorfs coming now" less semantically odd. Moreover, intentional explanations of the type, kimus flee because they believe that snorfs are coming, may seem less implausible.

I contend that TE-2 is not "accidental" in the sense intended by Pietroski; instead, it can appeal to the same common sense view of direct, perceptual cause that apparently motivates Pietroski's views. Nevertheless, the reader might find it disturbingly indirect: kimus are not repelled by the immigrant carnivores; they are, visually attracted to them. Jack and his descendants are spared the fate of most of Jack's contemporaries because Jack had the good fortune to have the right mutation. Still, many readers might find the claim that a B-token "signifies roughly, snorfs coming now" acceptable. But when a flock of mockingbirds passes by causing the kimus to flee, an intentional explanation of the sort, kimus flee because they believe snorfs are coming, seems decidedly odd. Perhaps we could diminish the sense of oddness by appealing to proper function, but then that of course is an appeal to Millikan's theory. Moreover, if we discover that snorf song is dialectical (and therefore, the version that causes Jack's B-token is in no sense essential to snorfness), and if we discover that kimus have no alternative means by which they can recognize snorfs as predators, we might be inclined to assent to the claim that "kimus wouldn't know a snorf from a hole in the wall."

I contend that TE-3, though accidental in Pietroski's sense, suggests that what matters most is not the common sense view of direct, perceptual cause, but instead whether or not our

intuitions are guided by a theory in light of which the co-variation can be interpreted as plausible (arguably this sense of plausibility contains a causal component). The crucial difference between TE-1 and TE-3 is that where TE-1 draws upon a biological concept, TE-3 draws upon a biblical one. I submit that an orthodox biologist would find the Millikan interpretation of B-tokens in TE-1 to be less semantically odd than the Millikan interpretation of B-tokens offered for Pietroski's thought experiment. Likewise, I submit that an orthodox Judaeo-Christian would find the Millikan interpretation of B-tokens in TE-3 to be less odd than the Millikan interpretation of B-tokens offered for Pietroski's thought experiment. And the difference in both cases is not the presupposed common sense view; rather it is whether or not intuitions are guided by a theory in light of which the co-variation can be seen to be plausible. It matters not whether the theory is the product of unexamined common sense, biology, or theology.

In sum, I contend that the sense of semantic oddness and intentional-explanation implausibility produced by Pietroski's thought experiment is not exclusively motivated by a reaction to Millikan's teleosemantics. Pietroski appears to be benefiting from an ambiguity: while it may well be that we prefer to believe that B-tokens are somehow directly grounded in their distal causes (it might be the case that our intuitions concerning direct grounding share a universal component), but all that is necessary is that we are able to see a connection, of some kind—just as an orthodox Judaeo-Christian might have no trouble seeing a connection between signs from god and safe havens. Moreover, I contend that TE-2 reveals problems inherent to Pietroski's common sense view.

To this point I take the arguments of section three to show the following: (1) Pietroski cannot avail himself of any standard, robust philosophical concept of "accident." They not only do not support his position, they call it into question. Therefore, he can only rely upon a common sense view. (2) The common sense view depends heavily upon inadequately explicated notions of directness, perception, and cause. Adequate explanation of perceptual reliability

might require recourse to an evolutionary argument, precisely of the type that Pietroski rejects and Millikan advocates. (3) Empirical theory shows that there might be legitimate reason to disallow thought experiments of the type employed by Pietroski (and here Millikan shares vulnerability). And, (4) thought experiments designed to be of a type with Pietroski's, through minor adjustments in background conditions, suggest that Pietroski may be helping himself to a previously hidden ambiguity and they reveal that semantic oddness and intentional-explanation implausibility also pose problems for thought experiments that are not P-accidental.

Pietroski is also concerned that Millikan fails to identify just what considerations or what findings might count against her theory. It can, however, be said on behalf of Millikan that though it is unlikely she or anyone could design a crucial experiment, the results of which could unequivocally be regarded as counting against her theory<sup>10</sup>, it is certainly possible to test "what-for"<sup>11</sup> questions of the type she asks "in the long run" and "against an ever-widening body of data" (Sober 1993: 128-129). Indeed this is standard practice for theories motivated by evolutionary/selectionist views.<sup>12</sup> As for a priori considerations, arguments given above suggest that at least some of the most critical worries advanced thus far can be vitiated.

Pietroski also worries that Millikan has not provided independent motivation for her view and that she risks allowing new (or otherwise undesigned) discriminatory abilities to drop out of intentional explanation--I will return to these at the end of section IV. But perhaps two points related to the issue of independent motivation can be raised at this juncture.

First, teleological accounts of the type provided by Millikan have already demonstrated their explanatory force in accounting for other traits of evolved organisms. Success in other,

---

<sup>10</sup> In fact on the Quine-Duhem Thesis, crucial experiments are ruled out. There may be no principled way to localize the bearing of evidence (e.g. Quine 1961).

<sup>11</sup> On Brandon's (1996: 41) account, what-for-questions and their answers are teleological in that for any trait A, "A's existence is explained in terms of effects of past instances of A; but not just any effects: we cite only those effects relevant to the adaptedness of possessors of A."

<sup>12</sup> Sober (1993: 27-28) shows, for example, how proceeding in this way led to rejection of the phrenologist's research program.

related domains,<sup>13</sup> is not necessarily transferable. But at least it is not self-evident that the mode of explaining why, say, fish have fins or humans have eyes, will not be up to the task of explaining why kimus have B-tokens; it is not obvious why mental representations should, in this regard, be exceptional. Empirical adequacy and explanatory force should be the criteria whereby teleosemantics is either found to be misguided or found to be appropriate. Second, where causal mechanisms are methodologically solipsist, both content and biological functions reach out beyond the confines of controlling causal processes. If we rest content on function, at least we are thereby able to "locate its point in a theoretical structure that inherently ... traffics in extrinsic relations to the environment."<sup>14</sup> This approach seems essential to any attempt at explaining representation-world relations.

#### SECTION IV

Discussion of Pietroski's worries continues here, but rather than continuing to discuss them in reference to his thought experiment, I here substitute a true case, a case that revolves about humans, not non-human animals. The case I describe is analogous to Pietroski's in all important respects, except where Pietroski dealt with what were apparently innate mental tokens, I here concern myself with learned mental tokens. This shifting of ground is consistent with Millikan's theory for she emphasizes that although design through natural selection may form the core of her paradigm cases, she allows for design through experience and learning (see section I).

Judging from Pietroski's choice of example I take it that he is primarily concerned with natural kinds. I follow his lead and give an example of what I take to be a natural kind and restrict most of my discussion to natural kinds.<sup>15</sup> Whether dealing with thought experiments or

---

<sup>13</sup> I here take it for granted that "success" is predicated on the realization that (a) sometimes a trait is not itself an adaptation but is a consequence of some other adaptation; (b) sometimes a trait is constrained by history; and, (c) sometimes a trait may have come about without in any sense having been selected. For discussion and examples, see Gould and Lewontin (1979). For criticism of the Gould-Lewontin position, see Dennett (1995: 267-282).

<sup>14</sup> See McGinn's (1989: 149-150) discussion of this point.

<sup>15</sup> For animal species there is considerable difference of opinion as to whether they are properly construed as natural

genuine empirical circumstances, concentrating on natural kinds allows us to say with a bit more confidence just what is relevant to our deliberations and what is a reasonable inference (cf. Wilkes 1988: 15). By concentrating on natural kinds the world is able to exercise a healthy constraining influence on our imagination.

I prefer actual cases because they come complete with background. This alone does not make it impossible for us to ignore relevant features that are troubling for our pet theory, but it may help make those features more difficult to ignore. Basing our arguments on "actual cases allows us to check our imagination against the facts, and our intuitions get strengthened and rendered more trustworthy" (Wilkes 1988: 48).<sup>16</sup>

The reason for shifting to humans is that discussion of intentionality among non-humans is even more problematic than it is for humans (e.g. Putnam 1992: 19-34).<sup>17</sup> At some point in the discussion of semantics it becomes necessary to confront the difference between having a true belief and having a successful belief. Humans (Putnam 1992: 30) are reflective creatures and are able to think about and criticize their practices from multiple points of view. Of a thought that I acted on I can both ask whether it satisfied my desires and whether it was true, and these are two distinct questions. For a dog, a cat, or a kimu, it is not at all clear that it makes sense to distinguish between true and successful beliefs. Furthermore, if the Deacon theory becomes better confirmed, it may be necessary to abandon the practice of conflating human and non-human intentionality in our critical reflections.

Ignaz Semmelweis discovered the cause of puerperal fever which was plaguing women who gave birth in the first maternity ward of the Vienna General Hospital from 1844 to 1848.<sup>18</sup> In

---

kinds, and if they are natural kinds, under what construal. For discussion, see Kornblith (1993: 50-52) and Sober (1993: 145-159). Concerning differences in how "species" is and has been used, see Keller and Lloyd (1992: 302-323). Also see Boyd's (1988) treatment of natural kinds as "homeostatic property clusters."

<sup>16</sup> For detailed discussion of the uses and abuses of thought experiments, see Brown 1991 and Sorensen 1992.

<sup>17</sup> Neither though are as problematic if one, merely for instrumental reasons, adopts Dennett's "intentional stance" (e.g. 1995). At minimum, it can sometimes confer on one substantial predictive power.

<sup>18</sup> I am drawing upon Hempel's (1966: 3-8) account of Semmelweis and Wingate's (1983: 363) description of

1844, as many as 260 of 3,157 mothers in the first ward (8.2%) died of the fever. These figures were especially alarming because in the adjacent second ward of the same hospital, which accommodated almost as many women, only 2.3% died of the fever. In telling the rest of the story I will adhere as closely as possible to the form of Pietroski's thought experiment.

The doctors had previously not been concerned to disinfect, even to wash, their hands after treating patients or performing autopsies. But, in virtue of a series of experiences--including witnessing the death by puerperal fever of a colleague whose finger had been punctured by a scalpel used in performing an autopsy, a scalpel covered with putrid matter, a type of matter whose odor could be removed through cleansing with chlorinated lime--Semmelweis came to produce tokens of a physically specifiable type B in the presence of chlorinated lime. Whenever Semmelweis performed an autopsy and his hands became covered with putrid matter, the container of chlorinated lime near the wash basin caused him to form a B-token. And Semmelweis (like his students) turned out to have a special "fondness" for chlorinated lime; i.e. other things being equal, Semmelweis would use chlorinated lime to wash away the putrid matter after it had been transferred to his hands during performance of an autopsy. So, after each autopsy, Semmelweis washed his hands and eliminated the putrid matter (in this way he stopped spreading infection to his patients). Learning took over (e.g. association of the odor and the fever, and hand-washing with elimination of the odor) and Semmelweis's token type proliferated among students and other doctors. There was no other reason for selection in favor of having that "chlorinated lime" token type.

Only later was it discovered that bacteria, especially streptococci, cause puerperal fever. The odor is not a property of the bacteria itself. Rather the odor is a consequence of festering that occurs after a person has been infected. The property--putridity--might best be described as a relational property of the bacteria and infected organism. It was also discovered that

---

puerperal fever.

numerous types of disinfectant or antiseptic can, with varying degrees of effectiveness and varying degrees of damage to human tissue, kill the bacteria.

Now compare the Semmelweis case to Pietroski's thought experiment. If I have succeeded in showing the two to be analogous in most relevant aspects, especially in how both mental tokens succeed in latching onto empirical regularities, then it should follow that (at the time) the content of a doctor's B-token signified roughly "less streptococci on hands washed in chlorinated lime." That anyway is what seems to follow from Millikan's teleosemantics. Even though streptococci don't cause (in the Pietroski sense) the doctor to think about and try to get rid of streptococci, and even though (to paraphrase Pietroski) the doctor wouldn't know a streptococcus from a hole in the wall; nevertheless, his B-token is, in an important sense, about streptococcus.

Just as in Pietroski's case, we have a constant conjunction; or, more precisely, within this corner of space-time, from the perspective of an organism concerned to decrease the likelihood of death, an organism that does not have microscopic vision but that is able to detect and learn to pay attention to a certain range of phenomena, a (near enough) constant conjunction holds. In hindsight (i.e. from the perspective of more penetrating scientific theories) we know that the constancy of the conjunction was only apparent; that the putrid matter caused no disease; that other solutions disinfect better than chlorinated lime; and, that that which caused death was not that which caused the B-token.

From the vantage of hindsight what can we say about the content of a B-token? Perhaps the most that can be said with confidence and without begging too many questions, is that if we consider the doctors' interests, it would court absurdity to insist that they were really interested in washing their hands with chlorinated lime solution. Washing with chlorinated lime was just a means of eliminating a form of matter, matter that afforded them indirect perceptual access to that which constituted the real object of their concern. To make explicit the analogy to Pietroski's thought experiment: the doctors can be regarded as kimus, the streptococci as snorfs.



To the extent that my arguments in section three succeed, we can see that it is not necessarily semantically odd to proclaim that B-tokens are about streptococci. Moreover, to the extent that we view mental representations as providing us with a means of tracking our interests in the world (c.f. van Frassen 1980: 40), it does not seem at all inappropriate to say that B-tokens are about streptococci.

It may be that for many people there will be a common sense interest in continuing to treat chlorinated lime, putrid matter, and disease as inherently linked. Perhaps a more accessible example of this would be the relationship that most people take to hold between upper respiratory viral infections, cold, damp weather, and going outside without a jacket. From a given vantage point, the conjunction may be just constant enough and just convenient enough (i.e. as bits of knowledge that are easily understood and put to practical use) to remain a matter of interest.

But as regards the scientific explanation of puerperal fever contagion, putridity drops out. After all, bacteria can be found almost anywhere, even on magnolia trees. As regards prevention, chlorinated lime is found to be inessential--both less effective and more harmful to tissue than other antiseptics.

When dealing with natural kinds--and I assume both bacteria and Pietroski's snorfs to be natural kinds (albeit the latter are a fictional version)--we are presumably concerned to carve nature at its joints. There may be many clues that lead to discovery and identification of natural kinds, but if these clues prove not to be essential properties, at least for those with the appropriate expertise, those clues/properties will not be regarded as part of the meaning.<sup>19</sup> It

---

<sup>19</sup> I don't mean to imply that this is the whole semantic story. For one thing, it cannot claim to be empirically adequate (cf. Searle 1992: 50). Millikan's view seems to work well for natural kinds and for representation-world relations. But, at minimum, I believe her view will have to be supplemented by something that explains our capacity for "fine-grainedness" of meaning, perhaps something along the lines of a more sophisticated conceptual role semantics (e.g. Block 1994). Lockwood (1989: 308-315) argues that even such a combination, though it may assist with understanding information-processing, would not be of much use to understanding meaning. Also note that Neander (1995) recommends restricting teleosemantics to conceptual "primitives". Concerning problems

matters for natural kinds how we link up to the world, but that link often will only be achieved through a complex pattern of indirect relations. The meaning of a natural kind doesn't change just because the manner in which we perceptually access the kind changes. And, recall from the discussion in section three, that *homo sapiens* at least may be predisposed to conceive of the world in terms of unobservable essences (viz. natural kinds).

And this suggests another possible independent motivation for favoring Millikan's view over a naïve causal view. Millikan's seems to suggest a better way of understanding our success at tracking our interests into areas where we are cut off from direct perceptual access. Our perceptual access to the world is both limited (e.g. the many things we can't see or smell) and prone to being distracted by information (e.g. bright colors and moving objects) that may handicap more than advance our ability to track our interests. As Millikan (1994: 251) observes of her interpretation of the magnetosome case: "what is represented by the magnetosome is not proximal but distal; no proximal stimulus is represented at all." At least where our interests are tied to natural kinds--e.g. bacteria, snorfs, or oxygen-free water--our semantic system may provide us with a means whereby we are enabled to escape perceptual limitations; it may lead us to look beyond proximal stimuli. And again, for this view, there is some measure of empirical support.

As for Pietroski's worry that on Millikan's account we risk allowing discriminatory abilities that are not the result of design to drop out of intentional explanations, it does clearly seem to be the case that at least sometimes new or otherwise unselected discriminatory abilities will figure importantly in explanations of behavior, sound, or motion.<sup>20</sup> What is not so clear is whether they will be needed for explanations of action, explanations that invoke intentional categories.

---

attendant to Neander's proposal, see Zawidzki (1999).

<sup>20</sup> See fn. 13.

For the record it should be noted that Millikan is not committed to the idea that all discriminatory abilities will necessarily fit into intentional explanation. Her theory clearly implies that there are no first-generation, direct proper functions. Accordingly, to become intentional, discriminatory abilities must undergo selection. As she (1993a: 82) observes: "Suppose that an inner representation were the sort of item that could be removed with tweezers and set under a microscope. Just as a beef heart that lies in the market is still a heart, the representation would still be a belief, desire, visual image or whatever it had been. Not that you could tell just by inspecting it and seeing what it could do, that it was this. You would have to know or guess its history."

If a kimus experiences a mutation that allows him to hear snorf song, a sound to which he is attracted, he will not be long for this earth. This unselected discriminatory ability would be important to explaining the kimus's death--but it is not at all clear that we would require an intentional explanation. Or, had Semmelweis been born with microscopic vision, vision that enabled him to see streptococci, though clearly this would affect his motions or behaviors (e.g. eating habits), it is not obvious how this would affect his actions--at least that is until after this ability had been fit into (via a selection process?) a body of beliefs and desires. After all, we already have perceptual access to an infinite variety of phenomena and much (perhaps most) of that to which we have access does not fit into intentional explanation of action--e.g. people who jump when they hear an explosion are not performing an action. "Most eye blinks and ducking reflexes, for example...are not intentional actions" (Millikan 1993b: 169). Pietroski might be onto something, but to make the point against Millikan he would need an additional argument that showed how the unselected discriminatory abilities would necessarily play a role in intentional explanation. Failing that, he would need to show, Millikan's denials aside, that her theory commits her to treating unselected discriminatory abilities as significantly contributing to intentional action.

But here the conceptual and empirical difficulties are complex. For example, it is not even self-evident just what constitutes unselected discriminating ability. Consider the *ey* gene in the fruitfly *Drosophila* (Dawkins 1997: 221-222); in normal circumstances it makes eyes. Because all of an animal's genes are present in all its cells—though only a minority are actually turned on in any given body part—cunning laboratory manipulation has made it possible to create fully formed, functional (at least they can be proven to be light-sensitive) eyes on the wings, legs, antennae, and elsewhere. What's more the genes (*small eye*) in mice and in humans (*aniridia*) that are essential to creation of normal eyes have DNA sequences almost identical to those found in the *Drosophila ey*. Here too clever manipulation has enabled researchers to introduce the mouse gene into *Drosophila* embryos. The result: the mouse gene turns on *Drosophila*'s eye-making machinery and creates not a mouse, but an insect eye. Similar DNA sequences have been found in molluscs, marine worms, and sea squirts. *Ey* may be universal among animals and it may be generally true that a version of the gene taken from a donor in one part of the animal kingdom can induce eyes in recipients from remote parts of the animal kingdom.

Knowing this is the case, what then should we say about, for example, a mouse gene turning on an ectopic human eye, an eye in a previously eyeless human body part? Does this count as having no history? As being unselected? This discriminating ability may not be typically turned on, but couldn't one argue that it was selected so as to be available, and that in its expressed form it inherits a derived status as selected? And certainly the history of such latent capacities is long.

Nevertheless, let's concede that the conceptual issues can be sorted out and that it can reasonably be regarded as unselected. The next step would be to determine whether or not its discriminatory abilities would sometimes be appropriately fitted into intentional explanations (as opposed to causal explanations of mere motion or noise). To do this one would need to check to see whether it performs in ways that typify intentionality. If we assume that current understanding of intentionality will hold into the future, by way of example, we can take a

criterion that Millikan (1993c: 109) recommends, the Brentano-criterion, the idea that intentional states can “be directed toward something that (doesn’t) exist” (e.g. a child’s beliefs about Santa Claus, or an adult’s beliefs about honest presidential candidates). As Millikan observes, “intentional” icons will sometimes mislead their users; they do this when there is no environmental feature onto which they can appropriately map. If we were to find an unselected discriminatory ability that can mislead in such a way, that can be “about” non-existents, we would have found a counter-example to Millikan’s theory, and Pietroski would be vindicated.

## SECTION V

Pietroski (1992: 278-281) contends that Millikan conflates teleological and intentional explanations. He claims that because both types of explanation sometimes make reference to the same properties, Millikan's teleological explanations of content may appear plausible. But, for Pietroski, even when, as with the case of frog representations of bugs crossing their field of vision, in which both types of explanation appeal to bugs, where teleological explanations show how a certain behavior (i.e. tongue-snapping) plays a useful role in the life of the system, intentional explanations show how tongue-snapping is “the product of how the system takes the world to be” (p. 278). Pietroski takes intentional states to have two distinct functions; Millikan regards Pietroski’s intentional explanations as merely a species of teleological explanations.

Pietroski develops another example to demonstrate that a single phenomenon can be explained in different ways. The phenomenon is the conservative slant of a newspaper; the categories invoked to explain this phenomenon are “ideological staff” and “powers-that-be.” He wants to allow both that writers write reflections of what they believe and that publishers and/or advertisers tend to hire those who agree with them and fire those who disagree. For Pietroski, the two forms of explanation are distinct and separable--though it is often the case that both are true. The “ideological staff” explanation is ahistorical; the “powers-that-be” explanation is historical. And Pietroski wants to emphasize that, whether the subject be frogs or newspapers,

"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" (p. 280).

Teleological explanation is of the first type; intentional explanation is of the second. To move from one form of explanation to the other is not to show that the need for one has been obviated; rather, it is "to change the subject" (p. 280). For Pietroski, Millikan's account of kimu B-tokens may be interesting in its own right, but it is not about intentionality. She has changed the subject. On Pietroski's account (p. 280), (1) teleological explanations are not in any sense "prior to" intentional explanations, (2) intentional explanations "often precede" teleological explanations, and (3) "at a minimum, intentional explanations are independent of ethological (teleological) explanations."

## SECTION VI

There are times when "changing the subject" does amount to obviating the need for one or more of the proposed explanations. Certain entities, laws, and models can be eliminated; some entities, for example, despite what people think them to be, fail to refer.<sup>21</sup> There is no longer any explanatory work to be performed by an entity like phlogiston, by the laws of phrenology, or by Ptolemaic models. With this much, doubtless Pietroski would agree. Of course he does not agree that there is no explanatory work left for his version of intentionality.

But consider a variant of his "ideological-staff" -- "powers-that-be" example. Suppose we find that whenever the "powers-that-be" change, say, from liberal to conservative, there is never any need to fire staff members. All or nearly all staff members, within a reasonably short period of time, reorient their ideology. Somehow (whether by rationalizing, compartmentalizing, becoming a neurotic, or however they manage it) they come to take the political world to be different than they had before. Before their editorials appeared to have been written by left-wing ideologues; later their editorials appear to be written by right-wing reactionaries.

---

<sup>21</sup> See Neander (1991: 172) and Churchland (1989: 12-17) for discussion.

If we found this to be a general condition, though we might still be able to put Pietroski's sense of intentional explanation to some use--e.g. to cover the interim period while everyone is busy reorienting--we would begin to suspect that one set of explanatory categories was leaning very heavily on the other. At a minimum, Pietroski's intentional explanations would not seem to be extremely independent. It may even begin to appear that "ideology" had little to do with explaining human actions. Perhaps, for example, ideology is so constituted that it allows for quick and facile adjustments in human commitments. It might then be that ideology's explanatory role is limited to after-the-fact rationalizations.

As for frog representations of bugs crossing their visual field or kimus B-tokens. These traits are innate; there is no room for ideological maneuvering. If a family of kimus settle in a neighborhood of red snorfs or a family of frogs settles in a swamp inhabited by poisonous flies, the B-tokens and the bug representations will disappear because their hosts will have been selected against. The bug- or B- token way of taking the world to be comes to an end. Here the categories of intentional explanation lean even more heavily than they do in my variant on the "ideological-staff"--"powers-that-be" example.

But the issues here rapidly become thorny, especially as they apply to ideology and human learning. Here Millikan's theory remains undeveloped, the Semmelweis case analyzed in section four above being one of the few attempted applications. And I have restricted myself to natural kinds like bacteria and animal species. It remains to be seen whether or not an empirically adequate, explanatory application can be made to political kinds (e.g. beliefs about the Three Principles of the People or Taiwan Independence) or to other non-natural kinds.

The Semmelweis case does though suggest that at least when reflecting on human learning a sharp historical-ahistorical distinction may be unwise. A dynamic, historically-embedded model—like Millikan's—may prove necessary to account for how, in the span of less than a decade, our representations of items in the world can change several times, honing in ever more narrowly on distinct items in the world. Further, Millikan's view seems especially amenable to

treatment of cases in which there are contemporaneous, competing representations. So perhaps it will so develop as to be able to handle political kinds.

## CONCLUSIONS

Pietroski's thought experiment can avail itself of no refined philosophical sense of "accident." In fact, philosophical concepts of "accident," if applied to Pietroski's thought experiment, all provide motivation for resisting his conclusions. The common sense notion of "accident" implied by Pietroski's critique has also been shown to be of dubious value. And the intuitions motivated by his thought experiment appear to be less the product of resistance to Millikan's theory than they are the product of failure to see in what light things, events, or processes could be thought about in the way Millikan proposes. My three counter-thought experiments help diminish the sense that Millikan's view results in semantically odd interpretations and that it leads to implausible intentional explanations.

I have also argued that: (1) at least as regards natural kinds and representation-world relations, Millikan's view can usefully apply to human learning. (2) Her view, just like other teleological approaches is open to refutation. (3) There is some degree of independent motivation for her view. (4) Her theory may require revisions, but whether or not there will be such a need is contingent upon conceptual refinements and empirical discoveries. I identify areas in which such refinements are arguably necessary, and such discoveries possible. And (5) I argue that Pietroski's call for a historical-ahistorical distinction might impede understanding of some aspects of human learning. Millikan's resistance to this distinction provides us with an appropriately dynamic theory.



## BIBLIOGRAPHY

Block, Ned

1994 Advertisement for a Semantics for Psychology. IN Stich and Warfield, Eds. *Mental Representation*. Oxford: Blackwell.

Boyd, Richard

1988 How to be a moral realist. IN G. Sayre-McCord, Ed. *Essays on Moral Realism*. Ithaca: Cornell University Press.

Brandon, Robert N.

1996 *Concepts and Methods in Evolutionary Biology*. NY: Cambridge University Press.

Brown, James Robert

1991 *The Laboratory of Mind: Thought Experiments in the Natural Sciences*. London: Routledge.

Calvin, William H.

1996 *How Brains Think: Evolving Intelligence, Then and Now*. New York: Basic Books.

Cartwright, Nancy

1983 *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Churchland, Paul M.

1989 Eliminative Materialism and the Propositional Attitudes. IN *A Neurocomputational Perspective*. Cambridge MIT Press.

Dawkins, Richard

1997 The eyeless gene. IN Ridley, Mark, Ed. *Evolution*. Oxford: Oxford University Press.

Deacon, Terrence

1997 *The Symbolic Species: The co-evolution of language and the human brain*. London: Penguin Books.

Dennett, Daniel

1993 Back from the Drawing Board. IN Bo Dahlbom, Ed. Dennett and His Critics. Oxford: Blackwell.

1995a Evolution, Error, and Intentionality. IN Moser and Trout, Eds. Contemporary Materialism. NY: Routledge.

1995b Darwin's Dangerous Idea: Evolution and the Meanings of Life. NY: Simon and Schuster.

Dretske, Fred

1994 Misrepresentation. IN Stich and Warfield, Eds. Mental Representation. Oxford: Blackwell.

Frassen, Bas van

1980 The Scientific Image. Oxford: Clarendon Press.

Gould, Stephen Jay and Richard Lewontin

1998 The spandrels of San Marco and the panglossian paradigm: a critique of the adaptationist 1998 programme. IN Ridley, Mark, Ed. Evolution. New York: Oxford University Press.

Gregory, Richard L.

1994 Even Odder Perceptions. New York: Routledge.

Hempel, Carl G.

1965 Aspects of Scientific Explanation. New York: The Free Press.

1966 Philosophy of Natural Science. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

1991 Laws and Their Role in Scientific Explanation. IN Boyd, Gasper, and Trout, Eds. The Philosophy of Science. Cambridge: MIT Press.

Inwagen, Peter van

1993 Metaphysics. Boulder: Westview Press.

Keller, Evelyn Fox and Elisabeth A. Lloyd, Eds.

A Defense of Millikan's Teleosemantics

1992 *Keywords in Evolutionary Biology*. Cambridge: Harvard University Press.

Kornblith, Hilary

1993 *Inductive Inference and Its Natural Ground*. Cambridge: MIT Press.

Lewis, David

1993 *Causation*. IN Sosa and Tooley, Eds. *Causation*. NY: Oxford University Press.

Lockwood, Michael

1989 *Mind, Brain, and Quantum*. Oxford: Blackwell

McGinn, Colin

1989 *Mental Content*. Oxford : Blackwell.

Millikan, Ruth Garrett

1989 In Defense of Proper Functions. *Philosophy of Science* 56: 288-302.

1993a *Thoughts Without Laws*. IN Ruth Millikan's, *White Queen Psychology and Other Essays for Alice*. MIT Press.

1993b *The Green Grass Growing All Around*. IN Ruth Millikan's, *White Queen Psychology and Other Essays for Alice*. MIT Press.

1993c *On Mentalese Orthography*. IN Ruth Millikan's, *White Queen Psychology and Other Essays for Alice*. MIT Press.

1994 *Biosemanantics*. IN Stich and Warfield, Eds. *Mental Representation*. Blackwell.

1995 *Explanations in Biopsychology*. IN Heil and Mele, Eds. *Mental Causation*. New York: Clarendon Press.

1999 *A Common Structure for Concepts of Individuals, Stuffs, and Real Kinds: More Mama, More Milk and More Mouse*. IN E. Margolis and S. Laurence, Eds. *Concepts: Core Readings*. MIT Press.

Neander, Karen

1991 *Functions as Selected Effects: The Conceptual Analysts's Defense*. *Philosophy of Science* 58: 168-184.

1995 Misrepresenting and Malfunctioning. *Philosophical Studies* 79:109-141.

Papineau, David

NY: 1990 Truth and Teleology. IN Dudley Knowles, Ed. *Explanation and its Limits*. Cambridge U. Press.

1993 *Philosophical Naturalism*. Oxford: Blackwell.

Pietroski, Paul M.

1992 Intentionality and Teleological Error. *Pacific Philosophical Quarterly* 73: 267-282.

Putnam, Hilary

1992 Does Evolution Explain Representation? IN *Renewing Philosophy*. Cambridge: Harvard U. Press.

Quine, W. V.

1961 *From a Logical Point of View*. Cambridge: Harvard U. Press.

1994 *Natural Kinds*. IN Hilary Kornblith, Ed. *Naturalizing Epistemology*, Second Edition. MIT Press.

Rey, Georges

1997 *Contemporary Philosophy of Mind*. Oxford: Blackwell.

Salmon, Wesley C.

*Causality and Explanation*. New York: Oxford University Press.

Searle, John

1992 *The Rediscovery of Mind*. Cambridge: MIT Press.

Sklar, Lawrence

1993 *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. NY: Cambridge University Press.

Sober, Elliott

1993 *Philosophy of Biology*. NY: Oxford U. Press.

Sorensen, Roy

A Defense of Millikan's Teleosemantics

1992 Thought Experiments. New York: Oxford Press.

Warfield, Ted A. and Stephen P. Stich

1994 Introduction. IN Stich and Warfield, Eds. Mental Representation. Oxford:Blackwell.

Wilkes, Kathleen

1988 Real People. Clarendon Press.

Wingate, Peter

1983 The Penguin Medical Encyclopedia, 2nd Ed. London: Penguin

Zawidzki, Tadeusz Wieslaw

1999 How to criticize Millikan: The Problem of Mythological Content.

Philosophy-Neuroscience-Psychology (at Washington University in St. Louis)

Technical Report 99-002.

# 為密勒坎（Millikan）的目的論語意學辯護

藍 亭

國立政治大學英語系副教授

（收稿日期：1998年8月7日；修訂日期：1999年4月30日；授受刊登日期：1999年9月27日）

## 摘 要

對密勒坎目的論觀點之語意學的主要批評是皮托司基（Pietroski）的指控：1.在密勒坎的理論中可能心靈表徵與引起它們的肇因完全不連貫；2.密勒坎沒有提出對其理論的否證條件；3.她的觀點極端且不能提供充分的動機；4.她忽略了未被選擇的分辨能力可能在意向性說明中扮演的角色；以及5.她把目的論及意向性的說明混為一談。

雖然皮托司基的批評，特別是第一項，直覺上有某種程度的說服力，但我論證它的實質內涵並不若表面所呈現的重要，且不能對密勒坎的語意學構成挑戰。此外，我嘗試說明密勒坎的理論不只適用於天擇的個案，也適用於人類的學習。再者，我提出了密勒坎的理論中，那些部分予以修改或許是適當的，而那些部分在我們達到某種概念上的澄清及更了解相關的經驗科學資料之前，是不能有所進展的。

**關鍵詞：**語意學，目的論，心靈表徵，天擇，說明