Understanding Intentionality in terms of Natural Selection: a Defence of Teleosemantics

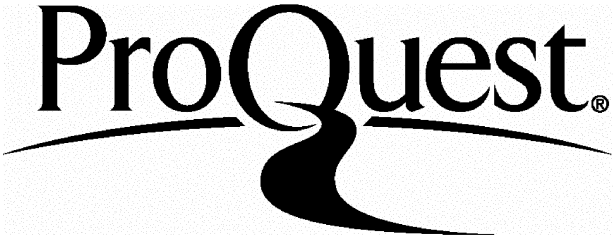
James Phillip Frank Tartaglia

Department of Philosophy

University College London

M.Phil. Thesis

1998



Understanding Intentionality in terms of Natural Selection: a Defence of Teleosemantics

By J.P.F. Tartaglia

Abstract

Rather than simply posit intentionality as a *sui generis* feature of the world, some philosophers attempted to analyse it. Chisholm appealed to logical features of the language used to describe mental states, to reformulate Brentano's thesis. Showing that intensional language is needed to describe mental states in order to preserve fine-grained meaning distinctions influenced the approach of trying to find a reduction of representational content. The approach of trying to reduce content seems to run counter to Brentano's contention that *all* mental phenomena are intentional, but teleosemantics does not rule out this possibility. And the contention that *only* mental phenomena are intentional is defensible.

Simple covariation cannot account for misrepresentation or robustness. Asymmetric dependence theory relieves these difficulties, but Fodor's conditions are not sufficient for representation, and without normativity in the account, the direction of the dependency relations has no explanation. Dretske introduces normativity via evolutionary, teleological considerations. But some misrepresentations have been selected for. Millikan's teleosemantics focuses on the evolutionary significance of representation consumption,

defining mental representations in terms of their historically determined proper functions as consumed. The theory can be seen as an extension of traditional functionalism.

Various objections that teleosemantics cannot make accurate content ascriptions, can all be answered satisfactorily from the resources of the theory. However, the theory has the counterintuitive consequence that a randomly generated physical duplicate of a thinker could not have intentional states. Upon closer inspection, this example turns out to be a vivid illustration of the thoroughly naturalistic conception of mind which the theory presents us with. Jackson claims that teleosemantics has the wrong theoretical status to answer important philosophical questions about content, but the theory has the right status to explain intentionality. Considerations about natural selection offer genuine insight into intentionality, showing how it fits into the natural world.

Contents

I - Intentionality: A Phenomenological Discovery

It was Franz Brentano, considered the founder of phenomenology, who brought intentionality within the ambit of twentieth century philosophy. He saw intentionality as the mark of the mental, the defining feature of mind which nothing non-mental could lay claim to. Edmund Husserl, Brentano's pupil, agreed that "all consciousness is 'intentional',"[1] and adopted the thesis as a central tenet of his own philosophy. The other great phenomenologists, Heidegger, Sartre and Merleau-Ponty, all followed Husserl in accepting Brentano's thesis, a notable fact considering the extent of their differences on other matters.[2]

Brentano himself held that when a subject's thought is directed upon external objects, an intentional relationship holds between the subject's mind, and the phenomena experienced by the subject. He makes this explicit in response to a commentator:

> What we think about is *the object* or *thing* and not the 'object of thought'. If, in our thought, we contemplate a horse, our thought has as its immanent object - not a 'contemplated horse', but a *horse*.[3]

---

[1] (Husserl 1929), p. 119

[2] Hence, Jean-Paul Sartre "It is of the very nature of consciousness to be intentional, and a consciousness that ceases to be a consciousness of something would *ipso facto* cease to exist." (Sartre 1940), p.211
And Maurice Merleau-Ponty, "... consciousness itself [is] a project of the world, meant for a world which it neither embraces nor possesses, but towards which it is perpetually directed." (Merleau-Ponty 1945), p. xvii
The situation with Martin Heidegger is more complicated for he thinks that "the intentionality of 'consciousness' is *grounded* in the ecstatical unity of Dasein" (Heidegger 1927), note xxiii, p. 498. For Heidegger, "intentionality is not an objective, extant relation between two things extant but, as the comportmental character of comporting, a determination of the subject." (Heidegger 1975), p. 61
Nevertheless, although Heidegger's conception of intentionality as an 'act-object' structure within Dasein's being-in-the-world, breaks with the tradition of Brentano and Husserl, Heidegger still clearly affirms Brentano's thesis: "Every comportment is a comporting-toward ..." (Heidegger 1975), p. 58

[3] (Brentano 1930), p. 77

Intentionality is seen as the way subjects are connected to what they are thinking about. Brentano understands intentionality as a relationship holding between mental states and the phenomena "intended" by those mental states, so that when somebody is thinking about the world, their thought is understood to be directed upon worldly phenomena.

For the phenomenologists, intentionality was a great discovery which could be put to use in explicating the world views they favoured. Intentionality has a key role to play within many of the classic texts in the idiom. Yet that tradition has little to tell us about the most puzzling aspect of intentionality: how is it possible, and what does it mean, for one thing to be about another? This is not an error of omission but rather a statement of policy, for intentionality was generally taken to be a basic feature of our existence. Merleau-Ponty neatly summed up this attitude when he wrote that,

> Our relationship to the world, as it is untiringly enunciated within us, is not a thing which can be any further classified by analysis; philosophy can only place it once more before our eyes and present it for our ratification.[4]

However, when the subject matter of intentionality was imported into the analytic tradition by Roderick Chisholm amongst others, further classification by analysis was exactly what was attempted. Rather than just describing and making use of the notion of intentionality, attempts were made to provide necessary and sufficient conditions for the application of the concept. Such an analysis was expected to explain away the puzzling

---

[4] (Merleau-Ponty 1945), p. xviii. Much the same point is made by Heidegger: "That the components of representing, judging, thinking, and willing are intentionally structured is not a proposition that can be noted and known so that, say, inferences can be made from it; rather, it is a directive to bring to mind what is meant by it, namely, the structure of comportments, and, by turning to the phenomena, to assure ourselves ever anew of the legitimacy of this assertion." (Heidegger 1975), pp. 65-66

aspects of intentionality by using less problematic concepts to provide a logical equivalent to the target concept of intentionality. We shall follow the important developments in the attempt to understand intentionality, from Chisholm onwards. In the process, the shape of the problem will be refined and developed, and the requirements of an adequate account of intentionality will be revealed. Then, the teleological theory of representation will be presented as the best candidate to meet those requirements.

II - Analysing Intentionality: Chisholm

Chisholm made one of the first important attempts to analyse intentionality,[5] taking Brentano's thesis as his starting point. Brentano claimed that every intentional act, "includes something as an object within itself".[6] By this, he meant that the object intended by an intentional act, exists within the intentional act itself. Brentano also talked about all and only psychological phenomena exhibiting "intentional in-existence"[7] to make the same point; the objects of intentional acts have in-existence in virtue of existing *in* the intentional act itself. Brentano found that the ability of mental phenomena to intentionally contain objects within themselves was an exclusive characteristic of the mental realm, and that, "No physical phenomenon manifests anything like it".[8]

Of central importance to Chisholm's understanding of intentionality was the idea that it cannot be inferred either that an object exists or does not exist, simply from the fact that the object is being thought about. Relations between physical objects on the other hand, do presuppose the existence of both of their relata; as Chisholm noted, "We can desire or think about horses that don't exist, but we can ride on only those that do."[9] Brentano originally placed no emphasis upon the ability of mental phenomena to

---

[5] However, Chisholm's was by no means the earliest attempt. In a 1921 discussion of Brentano's thesis, we find Bertrand Russell determined to further analyse intentionality: "The view here expressed, that relation to an object is an ultimate irreducible characteristic of mental phenomena, is one which I shall be concerned to combat." (Russell 1921), p. 15

[6] (Brentano 1874), p. 88

[7] ibid., p. 88

[8] ibid., p. 89

[9] Editor's introduction in (Chisholm 1960, ed.), p. 4

intentionally contain non-existent objects within themselves.[10] His was the wider point that any object of an intentional act is intentionally contained within the act itself.[11] However, the idea of the intentionality of the mental as the distinctive ability of mental states to be directed upon objects in such a way as not to presuppose that those objects exist, has been influential within the literature since Chisholm. Chisholm found what he called the "psychological attitudes"[12] to be most clearly illustrative of this notion of intentionality. Beliefs are paradigmatic psychological attitudes. To have a belief is to have a belief about something, and yet the object of belief need not exist, as is evidenced most clearly by beliefs about things like fairies and unicorns.

That beliefs are intentional, and thus always about something, is true whether the belief in question is conscious or unconscious. Somebody may not realise that they have a certain belief, but nevertheless, it is a condition of their unconscious state being a belief that it be about something. Somebody cannot have a belief if there is not something which they believe. The same goes for desires, hopes, expectations, fears, wishes, assumptions, and all those other types of mental state properly classified as psychological attitudes. The name fits these sorts of mental state because they all seem to be attitudes to something other than themselves. Having a belief is having a certain attitude to whatever it is that you believe. The attitude you have is a believing attitude. In the same way, having a desire is having a desiring attitude to whatever it is that you desire.

---

[10] However, this did become an issue for Brentano towards the end of his life; see the Appendix to (Brentano 1874), pp. 271-307, prepared for the 1911 edition of that work, in order to "explain and defend, as well as to correct and expand" upon the theory he had originally presented.

[11] This point is made in (Crane 1998).

[12] The following discussion of Chisholm is based upon (Chisholm 1957), chapter 11.

Chisholm found that the sorts of sentences used to describe psychological attitudes, "intentional sentences" as he called them, have three distinctive features. Firstly, they do not license existential generalisations, which is just to say that they use names and descriptions without necessarily implying that those names or descriptions refer to anything, so that consequently, you cannot always quantify into the naming or referring position of such sentences. Thus from an intentional sentence of the form "Fa", it cannot be inferred that "(∃x) Fx". By this criterion, "Diogenes looked for an honest man", which has the form "Rab", is an intentional sentence, since it does not have to be read as implying that there are any honest men, and so we cannot infer from "Rab" that "(∃x) Rax". Secondly, intentional sentences do not necessarily imply the truth or falsity of any propositional clauses which they contain. For example, "James believes there are tigers in India" does not imply that there are or are not any tigers in India. Thirdly, intentional sentences do not guarantee that truth will be preserved when substitution of co-referring terms occurs within them. For example, if it is true that "James knew in 1944 that Eisenhower was the one in command", it does not follow that it is true that "James knew in 1944 that the man who was to succeed Truman was the one in command", even though "Eisenhower" and "the man who was to succeed Truman" are co-referring terms.

Chisholm used these distinguishing features of the language used to describe psychological attitudes in order to reformulate Brentano's thesis. Brentano's claim that all and only mental states are intentional became the claim that intentional sentences are needed to describe the psychological attitudes, whereas they are not needed to describe non-psychological phenomena. Chisholm's reformulation brought the topic of

intentionality within the ambit of the linguistic philosophy of his day. However, despite analogies between Chisholm's claim that intentional sentences are needed to describe psychological attitudes, and Brentano's claim that intentionality uniquely characterises psychological phenomena, the two are not equivalent.[13]

The most obvious manifestation of this non-equivalence is that Chisholm claims only that intentional sentences are needed to describe psychological *attitudes*, but Brentano thought that *all* psychological phenomena exhibit intentionality. Brentano discusses the intentionality of sensations and emotions,[14] and so if Chisholm were right that intentional sentences are needed to describe intentional phenomena, then in reformulating Brentano's thesis, he ought to have said that intentional sentences are needed to describe all psychological phenomena, and not just psychological attitudes. He did not say this because it is false; "John's leg hurts" and "John is elated" are not intentional sentences on Chisholm's criteria. So whereas Brentano's thesis was that all and only mental phenomena are intentional, Chisholm's reformulation can at most claim that only mental phenomena are intentional.

For Chisholm's claim that only the psychological attitudes, and consequently only mental phenomena, need to be described with intentional sentences to be equivalent to the half of Brentano's thesis which claims that only mental phenomena are intentional, Chisholm needs the extra premise that intentional language is only needed to describe intentional phenomena. Chisholm never states this premise. Rather, in offering a linguistic reformulation of Brentano's thesis, he simply assumes that intentional sentences are only

---

[13] Chisholm does qualify his position: "We may now re-express Brentano's thesis - or a thesis resembling that of Brentano - by reference to intentional sentences." (Chisholm 1957), p. 172

[14] For example (Brentano 1874), pp. 78-79.

needed to describe intentional phenomena. This assumption is false. Sentences which are intentional by Chisholm's criteria are also needed to talk about modality and probability. Modal sentences like "Hague is necessarily a man" must be intentional sentences. For if "Hague is necessarily a man" allowed the truth-preserving substitution of co-referring terms, then since "Hague is the Leader of the Conservative Party" is true, we would be able to substitute "The Leader of the Conservative Party" for "Hague" into the sentence. But this yields the false sentence "The Leader of the Conservative Party is necessarily a man". Sentences about probability like "Hague will probably be the next Prime Minister" must also be intentional sentences. For suppose that Prob. (Hague is the next Prime Minister) = 0.25. Since Prob. (Hague is Hague) = 1, because Hague's self-identity is a necessary truth, then if sentences about probability allowed truth-preserving substitution of co-referring terms, we could substitute "the next Prime Minister" into "Prob. (Hague is Hague) = 1", which would entail the false conclusion that Prob. (Hague is the next Prime Minister) = 1.

Intentional sentences are needed to talk about modality and probability. In addition, they are also needed to report speech: it might be true that, "James said 'Hague is young'" but false that, "James said 'the Tory leader is young'". Since the fact that a phenomenon could only be described with intentional sentences does not mean that the phenomenon exhibits intentionality, then Chisholm's thesis that only the psychological attitudes need to be described with intentional sentences cannot amount to the claim that only mental states are intentional. Moreover, since intentional sentences must be used in all of the above contexts, then Chisholm's thesis must be false. So we must therefore

conclude both that Chisholm's thesis is false, and that its falsity has no bearing upon the question of whether or not Brentano's thesis is true, since all and only mental states might well be directed upon something, or about something, whether or not it is the case that intentional sentences are only needed to describe psychological attitudes.

The sentences which Chisholm called "intentional", John Searle would call "intensional".[15] Intensionality is a logical feature of some aspects of language. A sentence is intensional if it is non-extensional, where extensional sentences are those that license existential generalisations about the terms they contain, and the substitution of co-referring terms *salva veritate*. With hindsight, we might say that Chisholm tried to reformulate Brentano's thesis, which concerns the concept of intentionality, in terms of the concept of intensionality. Forging such an intimate link between the two concepts was a mistake according to Searle, not just because describing intentional phenomena is not the sole application of intensional language, but also because,

> They are not even remotely similar. Intentionality-with-a-t is that property of the mind (brain) by which it is possible to represent other things; intensionality-with-an-s is the failure of certain sentences, statements, etc., to satisfy certain logical tests for extensionality.[16]

For Searle, the concept of intentionality concerns mental states, and so should be analysed within the philosophy of mind. It would be misguided to focus an investigation into intentionality exclusively upon the language we use to talk about it, on Searle's view,

---

[15] (Searle 1983), pp. 22-26
[16] ibid., p. 24

because features of the language used to describe intentionality may not be features of intentionality itself, and *vice versa*.

III - An Intensional Circle

Chisholm thought that the way to show Brentano was wrong, was to find a way of describing the psychological attitudes without using what he called intentional language, and what we might call, following Searle, intensional language. Although the project of finding extensional language to describe the psychological attitudes, even if successful, need not show anything about intentionality, and certainly would not show that Brentano was wrong, it is still true that we use intensional language to describe the psychological attitudes. The half of Chisholm's thesis which says that we do not need to use intensional language to describe non-psychological phenomena is false. But the other half of Chisholm's thesis which says that we need to use intensional language to describe the psychological attitudes still might be true.

To see if intensional language could be replaced by extensional language, Chisholm looked to logical behaviourism. It was the contention of logical behaviourism that sentences about the mind could be reduced to sentences about behaviour and behavioural dispositions. But Chisholm found that attempts to replace intensional sentences about psychological attitudes with extensional sentences about behaviour were circular. Take for example the sentence, "James believes there are tigers in India". The behaviourist hypothesis is that such a sentence can be analysed in terms of hypothetical statements about what James will do under certain conditions. One such hypothetical statement is presumably that James would answer in the affirmative if asked whether there are tigers in India. However he would only answer this way if he wanted to tell the truth, and if he was

asked in a language which he understood. Here the circularity becomes evident, for wanting to tell the truth is another psychological attitude. And understanding the language must involve mental states other than the belief about tigers, since the same belief could be expressed in many different languages.

The circularity cannot be avoided by referring to non-linguistic behavioural dispositions either. A possible candidate for such a disposition is that James would be disposed to travel to India if given the opportunity. Yet this obviously requires reference to other psychological attitudes; James will only be so disposed if wants to see tigers, if he would not rather go to the zoo, if he is absolutely sure that there are tigers in India, and so on. Chisholm found that every attempt which had been made at replacing intensional language with behavioural dispositions expressed in extensional language was ultimately circular.[17] One reason this was of interest to Chisholm is that it meant intentionality, as he construed it, constituted a threat to behaviourism.[18] It seemed that talk about the mind could not be reduced to talk about behaviour and behavioural dispositions because of the intentionality of the mental.

The fact that the intensional language used to describe the psychological attitudes could not be replaced by extensional sentences, also raised difficult questions for realism. For the intensional language used to describe psychological attitudes revealed something troubling about intentionality; that it could not be a straight-forward relationship between mental states and objects in the world, since the relationship could hold between mental

---

[17] (Chisholm 1957), pp. 184-185

[18] In fact, Chisholm tentatively accepted Brentano's thesis, and rejected behaviourism, on the basis of the failure of behaviourist attempts to replace intentional with extensional language; see (Chisholm 1958), pp. 510-520.

states and non-existent objects, and because whether the relationship held between a mental state and an object might depend on how the object was described. For although the truth-conditions of extensional sentences are determined by the entities which they refer to, so that whether an extensional sentence like "James tamed the tiger" is true depends only upon a state of affairs involving James and the tiger, the same cannot be said of intensional sentences.

The truth-conditions of intensional sentences are not simply determined by the entities and states of affairs which they refer to, but also by the manner in which those entities and states of affairs are referred to. For example, if "James believes that he tamed Pegasus" is true, "James believes that he tamed the winged horse which sprung from the blood of Medusa" might yet be false, if James does not know that Pegasus is the winged horse which sprung from the blood of Medusa. And anyway, the truth-conditions of "James believes that he tamed Pegasus" cannot be determined by Pegasus, since there is no such entity. The need for intensional language to describe the psychological attitudes raises difficulties for realism, for it seems to show that worldly objects and states of affairs cannot be all that determines the truth conditions of such sentences.

One way to express the difference between extensional and intensional language is to say that intensional language is more fine-grained. The metaphor reflects the fact that the truth-conditions of intensional sentences depend upon sense as well as reference, to make Gottlob Frege's distinction.[19] How an object is described in an intensional sentence about a psychological attitude reflects the sense attached to the object by the subject of the attitude, or to use another Fregean term, the mode of presentation of the object. For

[19] (Frege 1892)

example, it is possible to believe that Hague is young without believing that the Tory leader is young, since a person might associate different senses with the terms "Hague" and "the Tory leader", even though both terms actually refer to the same man. As a consequence, the language used to report beliefs needs to be intensional in order to reflect the fine-grained nature of belief. The same could be said of the other psychological attitudes. Extensional sentences, which are not so fine-grained, are not concerned with the modes of presentation of the terms they contain, and so have truth-conditions solely dependent upon reference.

With this in mind, it seems that the connection between the intentionality of mental states, and the intensionality of reports about them, is not as inconsequential as Searle suggests. Intensional language is needed to describe psychological attitudes, for extensional language ignores fine-grained distinctions of sense. Chisholm's thesis might be no use as a reformulation of Brentano's thesis, but it does draw attention to the connection between intentionality and sense. The objects of intentional states are objects under a certain mode of presentation, which is just to say that a sense, or meaning, is attached to these objects which are thought about. Sentences about psychological attitudes are intensional because they incorporate the meaning as well as the reference of their terms. The psychological verb "believes" in a sentence like, "James believes that Hague is young" indicates that the following clause is not only concerned with reference. For although "Hague" in the sentence does refer to the man Hague, it refers to him in a certain way, that is, as thought of by James. Chisholm's thesis that intensional language is

needed to describe psychological attitudes can be understood as the thesis that the psychological attitudes are irreducibly semantic.

W.V.O. Quine did not miss this moral in Chisholm's work. For Quine, the need for intensional language, or "intentional idioms", when describing the psychological attitudes, presented an ontological indictment of the psychological attitudes, just because it showed them to be irreducibly semantic. This can be seen from the following passage, where he is content to simply equate intention, understood as what an intentional state is about, with meaning:

> Brentano's thesis of the irreducibility of intentional idioms is of a piece with the thesis of indeterminacy of translation. One may accept the Brentano thesis either as showing the indispensability of intentional idioms and the importance of an autonomous science of intention or as showing the baselessness of intentional idioms and the emptiness of the science of intention. My attitude, unlike Brentano's, is the second.[20]

Quine's thesis of the indeterminacy of translation was an argument to show that since physical facts will not determine that two sentences are semantically synonymous, there can be no objective fact of the matter as to what a sentence means.[21] The point was not epistemological, but ontological; it was not that we cannot know what is meant by an utterance of a sentence, but rather that physical facts will not determine the correctness of a semantic claim. In the above quotation then, Quine is saying that if sentences about psychological attitudes cannot be replaced with extensional sentences, then physical facts cannot uniquely determine what psychological attitudes mean, or are about.

---

[20] (Quine 1960), p.221

[21] Except in the case of observation sentences.

For Quine, this meant that intentionality is incompatible with physicalism. Psychological attitudes could not be reduced to the extensional language of physical science, as their criteria of identity were indeterminate. So Quine introduced his "Double Standard".[22] This was the idea that the intentional idioms should be considered as a loose means of expression, useful for all purposes except, "limning the true and ultimate structure of reality".[23] Then, when ontology is to the fore, talk of the physical constitution of organisms and their behavioural dispositions should be substituted for talk of psychological attitudes. This would not preserve the meaning of psychological attitude talk of course; Quine accepted that intentional idioms were irreducible. Here he followed Chisholm, who had concluded that behaviourist attempts at reduction were always circular; that in trying to give a reductive analysis of one intentional idiom, there was always a need to introduce another. Quine understood this as a symptom of his thesis of indeterminacy; an indeterminate category could not be reduced to a determinate one. But on Quine's picture, the fact that a physical and behavioural gloss of intentional idioms would not preserve their meaning was not a problem, since he thought that there was no determinate meaning to preserve. Physical and behavioural descriptions of mental states would preserve everything that was literally true of them.

Quine's position can be seen as a natural extension of Chisholm's linguistic construal of intentionality. Chisholm saw that the intentionality of the mental conflicted with behaviourism because sentences about intentional states were non-extensional. But if the need for intensional language was what caused the conflict, then anything else needing

---

[22] ibid., pp. 216-221
[23] ibid., p. 221

intensional expression ought to raise the very same problem. In accepting that the need for intensional language to describe certain phenomena conflicted with behaviourism, as well as with physicalism and the logical constraints which he placed upon ontological discourse, Quine turned Chisholm's thesis on its head. For when Chisholm said that only mental states need to be described with intensional language, he meant simply that only the psychological attitudes needed intensional expression. But Quine on the other hand, adopted the position that anything needing intensional expression must be a mental phenomenon of some sort, even if we would not ordinarily suppose it to be. This was because he saw intensional language as the language of meaning, and meaning as irreducibly mental. Thus he applied his Double Standard to deflate the significance of intensional language, and safeguard physicalism.

Quine's position rests upon his indeterminacy thesis, and Chisholm's linguistic construal of intentionality. The former is notoriously controversial, and as has been shown, there are many reasons to be suspicious of the latter. Moreover, Quine's position might even be portrayed as a *reductio ad absurdum* of these two positions, since it leads to the rejection of something which common-sense takes to be obviously true, namely that we can have beliefs and desires with determinate contents. Of course, Quine might have been right, in which case an attempt to gain a greater understanding of intentionality would be fundamentally misguided. But if we are inclined to reject Quine's position, then it should now be clear that Chisholm's linguistic construal of intentionality leads to a dead-end. The insight this approach yielded was that the language needed to describe intentional mental states must be capable of making fine-grained distinctions of meaning; but intensional

language could not be reduced, and so further insight into intentionality was not available via that route. Far more progress has been made by concentrating on mental states themselves, rather than the language used to describe them. So we shall now turn to the attempts which have been made to understand intentionality through the production of theories of mental states.

IV - From Propositions to Representations

The need to use intensional language in order to preserve the fine-grained distinctions between psychological attitudes can be explained by thinking of these mental states as attitudes to Fregean propositions, since Fregean propositions are senses.[24] On this proposal, psychological attitudes would be more accurately described as propositional attitudes. The idea is that propositional attitudes involve an attitude to a state of affairs as presented in a certain way, rather than simply to a state of affairs. How a state of affairs is referred to or presented within a sentence can change the meaning which the sentence is used to express, which is just to say that it can make a difference in determining which proposition is expressed by the sentence, since Fregean propositions are to be understood as pure meanings. For instance, the sentence "Hague is young" can be used to mean something different to the sentence "The Tory leader is young". The two sentences can express different propositions even though, as uttered now, they are about the same state of affairs, in one sense of "state of affairs". Although both of these sentences are extensional, and so whether they are true or not is determined by their reference only, the situation is different with an intensional sentence like "James believes that the Tory leader is young". This sentence is intensional because its propositional clause, "the Tory leader is young", is about a proposition which James believes. If we substitute "Hague" for "the Tory leader" into the clause, then it will be about a different proposition, which James may

[24] As opposed to Russellian propositions, which are constituted by items at the level of reference, and hence are extensional.

or may not believe. The truth-conditions of the sentence depend upon both reference and meaning.

The presently salient difference between "The Tory leader is young" and "James believes that the Tory leader is young" is that the first sentence is about the Tory leader, but the second sentence is about James' belief about the Tory leader. As such, the truth-conditions of the first sentence are determined by the features of the objects or states of affairs referred to by the sentence, whereas the truth-conditions of the second sentence are determined by the features of James' belief referred to by the sentence, which means that the truth-conditions of the second sentence are determined by the mode of presentation of the objects and states of affairs which James' belief is about. In other words, whereas the truth-conditions of the first sentence depend upon the Tory leader, those of the second depend upon how the Tory leader is represented by James.

The present proposal, then, is that propositional attitudes like beliefs and desires are attitudes to Fregean propositions. This simply rearranges the problem of explaining intentionality, for we now need to know how these propositions relate to the world, whereas before, we were faced with the more vague question of how thoughts relate to their objects. However, the introduction of propositions introduces more questions to be answered, since Fregean propositions are abstract objects, and we have no account of how a relation to an abstract object might have effects on the behaviour of a subject of a propositional attitude. Neither do we have any account of what it means to grasp a proposition, for to have a propositional attitude such as a belief that Hague is young, it is clearly necessary to grasp the proposition that Hague is young.

These problems are resolved by thinking of propositional attitudes in terms of mental representations, internal states of a subject which it may be possible to construe naturalistically. Instead of talking about subjects having a certain attitude to a proposition which they grasp, we can talk about the subject having, or instantiating, a mental representation with a certain propositional content. On the mental representation picture, believing, desiring and hoping are relational properties which relate subjects to mental representations. Yet we talk about true beliefs, satisfied desires, and hopes which have been fulfilled, and such phrases attribute semantic properties to propositional attitudes as a whole. We can explain this by thinking of propositional attitudes as having their semantic properties in virtue of being relations to meanings.[25] If what a mental representation means is true, that is, if its truth conditions are fulfilled, then to have a believing attitude towards the propositional content of that representation, is to have a true belief. Likewise, to have a desiring attitude towards that content would be to have a satisfied desire. The semantic properties of mental representations are, as it were, superimposed upon propositional attitudes, so that the truth-conditions of mental representations become the satisfaction-conditions of propositional attitudes.

The idea that propositional attitudes are constituted by an attitudinal relation and a mental representation amounts to what Jerry Fodor calls "RTM": the Representational Theory of Mind.[26] Alternatives are possible, but this model has dominated recent work in philosophy of mind. Two sorts of relations are posited by RTM: relationships between subjects and mental representations, and relationships between mental representations and

---

[25] This position is presented in (Fodor 1990), chapter 1.
[26] ibid., p. 5

what they represent. An analysis of the first sort of relationship would be an analysis of attitudes like believing and desiring. An analysis of the second sort of relationship would be an analysis of intentional content. On the RTM model, the task of giving an analysis of intentional content becomes the task of giving an account of the semantics of mental representation. To give an account of the semantics of a mental representation would be to say what it is for that mental representation to mean what it does, to represent what it does, or to adopt the standard terminology, to have the content that it does, the content of a mental representation being that which it represents. An account of representational content would say what it means for mental states to be about whatever they are about. It would provide an explanation of the relationship between mind and world which we naturally suppose intentionality to be.

The task of analysing intentionality was converted into the task of producing a theory of representational content. The teleological theory which will ultimately be defended is one such theory of content. Such a theory is needed for two principal reasons, both of which have been recurring themes in the story so far. The first reason that a theory of mental content is needed, stems from the fact that intentionality and meaning are mental concepts. Philosophy of mind this century has been pre-occupied with the mind-body problem, which is the problem of how to reconcile the findings of physical science with the mental ontology presupposed in the history of philosophy, and suggested by ordinary language. Widespread commitment to physicalism, the doctrine that philosophy should defer its ontological commitments to the findings of physical science, and more recently naturalism, the doctrine that the philosophy should defer its ontological commitments to

natural science, thus incorporating the special sciences as well as physics, has made a problem of the fact that intentionality is mental. A theory of mental content should show how intentionality fits into the natural, physical world. The second reason that a theory of mental content is needed is to explain what it means for something to be about something else. Some sort of account is needed if intentionality is not just to be a mysterious relationship holding between mental states and the world, a stipulated property of mental things which resists further description.

V - All Mental Phenomena?

The idea that the correct theory of content would adequately and completely explain intentionality is *prima facie* at odds with Brentano's thesis. For according to Brentano, intentionality is a characteristic of all and only mental states. The task of explaining content is only meant to tell us about the intentionality of propositional attitudes, and not about the intentionality of any other sort of mental state. Concentrating on the attitudes to the exclusion of any other type of mental state is just what Chisholm was criticised for earlier.

The question of whether all mental states are intentional is important, for we need to know which types of mental states have this feature of intentionality which we are trying to explain. If Brentano was right that all mental states are intentional, then an account that only deals with propositional content would be too restricted. However, mental states other than the propositional attitudes, which can be broadly classified into two groups, sensations and emotions, are not obviously intentional. A sensation such as a pain is not obviously about anything other than itself. We might naturally suppose that to be in a mental state of being in pain is just to have a certain feeling. Likewise, we might naturally suppose that to be in an anxious state of mind is just to have a certain emotional feeling. Searle accepts this, and restricts his account of intentionality to the psychological attitudes.[27]

Recently however, some philosophers have made moves to defend Brentano's contention that all mental states, and not just the propositional attitudes, exhibit

---

[27] (Searle 1983), p. 1

intentionality. One such philosopher is Tim Crane,[28] who thinks that sensations and emotions can properly be called intentional. He thinks that pain is to be explained along perceptual lines. To be in pain is to be in a mental state which is directed upon part of the subject's body, and which predicates the sensation of pain of that part of the body. Crane mentions the fact that when we have a pain in the hand, we can concentrate on it by focusing our attention onto the hand, or try to ignore the pain by thinking of something else. This suggests that the pain is not just a feeling, but rather an awareness of a feeling in a part of the body. So we might say that sensation is perceptual awareness of one's body, and that pains and other sensations are intentional in that they are about our bodies.

Emotions are also intentional on Crane's account. He suggests that they might be approached as attitudes to oneself and one's position in the world, so that being anxious for instance, might be explained as a mental state about how potentially disturbing the world seems to the subject, and how precarious the subject's place within that world seems to be. Crane picks on "undirected anxiety", as this is an emotion which Searle took as paradigmatic of non-intentional mental states. This sort of anxiety, a concept appropriated from Kierkegaard, plays a key role in Heidegger's philosophy, where in affirming that it is indeed an intentional state, he proposed much the same explanation of it as Crane does. Heidegger sees this type of anxiety as fear which is not about anything in the world, but rather about the subject's way of existing in the world, which involves the subject's projection of itself into that world: "That in the face of which one has anxiety is Being-in-the-world as such."[29] Heidegger's position here is an affirmation of the subject's

---

[28] (Crane 1998). Another philosopher who takes this position is Michael Tye; see (Tye 1995), chapter 4.
[29] (Heidegger 1927), p.230

connection to the world; even the most general intentional state there is, which is not about any particular aspect of Being, is still about something. It is about Being as a whole, and that is because human Being is Being-in-the-world.

Crane does not need to accept this or any similar position about humans' being necessarily situated. All that he needs is the phenomenological evidence which supports such views. It does indeed seem that all of our mental states are about something or another. It is hard to imagine what it might mean to think without thinking about something. Emotions do alter the way the world seems to us, and so are about how the world seems. Sensations do seem to be about our bodies, even though expressing the idea this way sounds awkward. An objection might be made, that the feelings or qualia associated with mental states do not come to us as referring to something outside themselves. But qualia are only supposed to be aspects of mental states, the feelings which tell us which mental state we are supposed to be in.

So there may be good reasons to tentatively accept that all mental states are intrinsically intentional. What difference will this make to our investigation into intentionality? It will mean looking for a reduction capable of dealing with more than just the psychological attitudes. The teleological theory that will be defended offers just such a reduction. Most alternative reductions of the intentionality of the mental are clearly incompatible with an acceptance of Brentano's thesis, since they are just not equipped to explain the intentionality of mental states other than psychological attitudes. The existence of a viable contemporary defence of Brentano's thesis, counts strongly in favour of

reducing intentionality in a way that does not rule out the possibility of Brentano's thesis being true. It will be argued that the teleological theory does just this.

On the negative side, such a general conception of intentionality runs the risk of generating too general an account, which misses the real phenomenon. Sellars calls the intentionality of sensations "pseudo-intentionality",[30] because he thinks that our sensations only seem to be about something as a result of our illegitimately superimposing the genuine intentionality of linguistic representations onto them. If this were true, or if some emotional states really are undirected, then it would be a mistake to favour an account of intentionality which encompassed all mental states. So perhaps it is best to stick to the traditional programme of looking only for a reduction of the psychological attitudes. As it turns out, this is all well and good, since the teleological theory will be presented as the natural consequence of the failure of previous accounts to reduce the intentionality of the propositional attitudes. In accomplishing the reduction for the propositional attitudes, it will be argued, a theory was produced which can cope with the intentionality of other types of mental state.

[30] "Being and Being Known" in (Sellars 1963)

VI - Only Mental Phenomena?

The other side of Brentano's thesis, the contention that only mental phenomena exhibit intentionality, also bears upon the task of searching for a theory of content to explain intentionality. For if intentionality characterises other phenomena, then an account of content need not have explanatory priority. It might be that to explain intentionality, we should seek an account of some more fundamental phenomenon than mental content. It is to this issue that we shall now turn.

H.P. Grice's notion of "natural meaning"[31] seems to challenge the position that only mental states are intentional. We say that certain clouds mean rain, and that certain spots mean measles. So in a sense, clouds and measles indicate something, or are about something. It would seem that natural, non-mental phenomena can exhibit intentionality. Fred Dretske has given the fullest development of the notion of natural meaning.[32] Dretske points out that when there is nomic dependence between two events, then according to the mathematical theory of communication which he appeals to, we can speak of one event carrying information about the other. The mathematical theory of communication, in measuring the amount of information passing between sources, treats information as an objective commodity. Dretske adopts this tactic, and thereby develops an information-based semantic theory. For Dretske, information is a product of any causal covariation between types of events, so that on his picture, the natural meaning of clouds and spots is the information they carry about the rain and the measles which they covary with.

---

[31] (Grice 1957)
[32] (Dretske 1981, 1983)

The common objection to this proposal is that it fails to make Searle's distinction between intrinsic and derived intentionality.[33] Intrinsic intentionality is to be understood as a characteristic of mental phenomena only. Derived intentionality on the other hand, is imposed upon natural phenomena by subjects who have mental states which exhibit intrinsic intentionality. The idea is that clouds only mean rain in the sense that somebody who sees the clouds would then expect rain. Non-mental phenomena only mean something as interpreted by somebody, but mental phenomena are intrinsically intentional. This proposal is supported by remembering that although it is natural to talk about clouds meaning rain, there are many other phenomena for which it is not natural to talk about them meaning whatever they covary with. For instance, it would be odd to say that a falling tree means the sound it will make when it hits the floor. But that is what an account like Dretske's is committed to saying.

Dretske's position is quite the opposite of Searle's, as is clearly expressed by his motto: "In the beginning there was information."[34] For Dretske, the world was never semantically inert. Clouds meant rain before humans realised that they did; they must have done, for it is not as if our interpretation affects the clouds in any way. Natural meaning is, on his view, an observer-independent feature of the world. If we accept this, then it might be thought that we could substitute an investigation into the intentionality of the mental, for an investigation into the wider phenomenon of information. Mental states would be intentional in virtue of carrying information, Brentano's thesis would be false, and the production of a theory of mental content would not necessarily be the best approach to

---

[33] (Searle 1983), pp. vii-viii, and pp. 27-28
[34] (Dretske 1981), p. vii

answering questions about intentionality, since intentionality would not be an exclusively mental phenomenon.

But we do not have to decide between Searle and Dretske's attitudes to natural meaning, since Dretske's position does not jeopardise the significance of a theory of mental content for explaining intentionality. Dretske does not think that natural meaning is "derived" from the intentionality of the mental, but neither does he think that the intentionality of the mental can be explained solely in terms of natural meaning. The content of a mental representation, for Dretske, is determined by what information it was designed to carry.[35] The teleological notion of design is essential to Dretske's account. The reason for this is that natural meaning alone cannot account for the possibility of misrepresentation. Clouds cannot mean that it will rain if it will not rain. This contrasts with mental representation, since whether the content of James' belief means that it will rain, does not depend on whether it actually will rain. Beliefs can misrepresent the world, but natural meanings cannot. By appealing to teleology, Dretske accounts for misrepresentation as what happens when a mental representation does not carry the information that it is its teleological function to carry.

Natural meaning might be given a key role in a theory of intentionality, but it will not in itself explain what intentionality is without the addition of some extra factor, such as the notion of teleological function which Dretske uses for his information-based theory of content. Consequently, which position to take on natural meaning is not a question which makes any difference to whether we should look for a theory of mental content to explain intentionality. It should only affect which theory of mental content we choose; those who

---

[35] (Dretske 1986)

view natural meaning as exhibiting only derived intentionality will not tend to favour an information-based theory.

The other reason to question whether only mental states are intentional is that we ordinarily suppose language to be a source of meaning. Sentences can express propositions, and so can be about objects and states of affairs. However, it is natural to suppose that words only have meanings as used by people who understand them. On this view, the sounds and shapes of language have no meaning themselves; it is just that we associate certain meanings with those sounds and shapes. In which case, we could make Searle's distinction, and say that linguistic intentionality is a derived intentionality. This is generally done; the view that meaning in language derives from meaning in mind is widely held. Chisholm held it, comparing the way sentences derive their intentionality from thought to the way the moon derives its light from the sun.[36] And Searle employs the view as a bedrock for his work on intentionality:

> ... philosophy of language is a branch of philosophy of mind. The capacity of speech acts to represent objects and states of affairs in the world is an extension of the more biologically fundamental capacities of the mind (or brain) to relate the organism to the world ...[37]

But the view Searle articulates above, although very plausible, is not uncontroversial. It assumes that mental states rather than language have an intrinsic intentionality which can be explained without appeal to any other sort of intentionality. On this view, language must be interpreted in order to represent something, but thought needs

---

[36] In a letter to Wilfrid Sellars, published in (Chisholm 1958), p. 533.
[37] (Searle 1983), p. vii

no interpretation in order to be representational. However, some philosophers have thought it a mistake to think of mental phenomena as intrinsically intentional, and hence in a different situation to non-mental phenomena. This view takes the explanation of the intentionality of thought as on the same footing as the explanation of the intentionality of language. The idea is that both are to be explained in terms of the rules governing their use. If we are inclined to think that all thought is linguistic, then accounting for the intentionality of language and for the intentionality of thought should be closely related enterprises. Donald Davidson,[38] who holds the view that a being cannot be a thinker unless it has a public language, thinks that explaining the intentionality of thought and language would be a simultaneous exercise.

Another philosopher with a similar neo-Wittgensteinian view on this subject is Wilfrid Sellars.[39] He thought that a theory of the use of language is all that is needed to account for intentionality. Sellars' "PMese" theory, as he called it, was based on the approach he saw taken in *Principia Mathematica*, which was to allow rules of inference to determine the meaning of terms, rather than allowing the meanings of terms to determine which rules of inference are classified as valid. On this view, for a symbol to have a meaning is just for it to be caught up within a system of rules which operate over the syntactical form of that symbol. Language use is such a system of rules, and so a theory of the use of words would tell us how a word could have a meaning, or be about something other than itself. The intentionality of thought, on Sellars' view, was a result of mental states having the same functional roles as linguistic items, since we think in a language.

---

[38] (Davidson 1974)
[39] (Sellars 1958, 1963, 1981)

Arguably, these sorts of view have been overtaken by the turn of events in recent philosophy of mind. For the search for a theory of mental content is the search for a non-circular account of the semanticity of mental states. Such an account would say what it is for a mental state to be intrinsically intentional, and would thereby explain why mental states do not need to be interpreted. To say that mental states are intrinsically intentional, in itself explains nothing, and so it is understandable that philosophers like Sellars were not satisfied. For it is not enough to simply stipulate that certain entities, mental states, are meaningful, without giving any explanation of what that meaning consists in. But the project of finding a theory of mental content, has the very aim of finding a theory which will explain what the intentionality of the mental consists in.

Philosophers of language can introduce semantics into their theories of public language by making use of the intentionality of the mental. Grice takes this approach and analyses speaker-meaning in terms of, amongst other things, the intention a speaker has to get the person hearing their utterance to form a certain type of belief on the basis of it.[40] The idea that language and thought might be governed by the same, or by a similar systems of rules, is not usually taken as an argument for explaining the intentionality of thought in terms of the intentionality of language, these days. Rather, it tends to be treated as evidence in favour of Jerry Fodor's "Language of Thought" hypothesis,[41] the idea that propositional attitudes have structural properties akin to public language expressions. Since this hypothesis only concerns the syntax of propositional attitudes, it does not affect the need for a theory of mental content. The idea that a use theory of meaning for public

---

[40] (Grice 1989)
[41] (Fodor 1975), and the appendix to (Fodor 1987), pp. 135-154

languages would explain intrinsic intentionality, is not a live option. But that is not to say that the rules governing transitions between a subject's mental representations, which might parallel rules governing the use of public language, might not have a prominent part to play within a theory of mental content. It is the contention of Functional Role Semantics, which will be discussed later, that they do.

VII - Covariation

The aim of this essay is to defend the teleological theory of representation as giving the correct account of intentionality. As such, the concern up until now has been to show why we need an account of intentionality, and to show that a theory of mental content is the right sort of account to give. The next principal concern of the essay will be that of giving an exposition and defence of the teleological theory, and then of showing the consequences of accepting the theory for our understanding of intentionality. In the interim, we will talk about some of the alternative approaches to representation which have been taken. Our concern will be with causal approaches to reducing content, where by causal, we mean any account which defines content in terms of causal interactions between organisms and their environments, so that covariation and teleological theories both count as causal. Other theories will be treated briefly because the aim of this essay is not to give a detailed survey of the advantages and disadvantages of all the various positions that can be held in the field of mental representation. Rather, it is to defend one such position. Consequently, only so much will be said about other theories as is necessary to reveal the central difficulties facing them. Then the teleological theory will be presented as providing a natural solution to such difficulties. By juxtaposing the teleological theory against its related rivals, its advantages should be made vivid.

To provide a non-circular definition of representational content, a theory needs to say what "X represents Y" is equivalent to, for any mental representation X, without making use of any intentional or representation notions. In doing so, such a theory of

mental content would explain what intentionality is, since intentionality is the relational property holding between mental representations and the conditions that they represent. Recent attempts to provide just such a reductive account fall broadly into two camps; those which explain mental representation in terms of the cause of the intentional state, understood as the conditions under which representations are produced, and those that concentrate on the effects the representation has, as consumed by the thinker. Covariation theories, as propounded by Jerry Fodor and Fred Dretske, belong to the former, production-based camp, and Ruth Millikan's teleosemantics[42] belongs to the latter, for its focus is upon the consumption of representations.

The basic idea behind covariation is that representations are natural signs produced in response to, and bearing natural information corresponding to conditions in the world. What a mental representation is about is explained in terms of what caused the representation to be produced. Intentionality is seen as the upshot of the responses of organisms to their environments. So a mental state type X represents Y just if X is a reliable indicator of Y, where indication is understood causally. X reliably indicates Y because Y systematically causes X, and so, should Y vary, X will covary with it. Thus the content of an intentional state is, with refinements, identified with its cause. The crucial refinement centres around the concept of reliability. Reliable indication is taken as natural and not conventional causal correlation between cause and representation, so that a thought about mushrooms induced by the sight of a certain commercial logo is not classified as a representation of the logo by the covariation theory, since the relationship

[42] In this essay, the neologism "teleosemantics" will only be used in reference to Millikan's teleological theory of representation.

between the logo and the representation is not subsumed under a natural causal relation. Representation is a matter of non-arbitrary covariation, so that the relation between a representation and what it represents is a naturalistic causal relation.

Simple covariation faces two main problems: it cannot account for misrepresentation, or for what Fodor calls the "robustness" of meaning. The problem of accounting for misrepresentation is usually regarded as the greatest difficulty facing any theory of representation. There are two types of case. There are cases of misrepresentation in which the subject represents something as present when it is not. That mushrooms are the content of Jayne's mental state X does not guarantee that mushrooms will be indicated whenever she tokens X, for the cause of X might be abnormal, as is the case in hallucination. And there are cases of misrepresentation in which the subject represents what is present as something else. Such cases generate the problem of disjunctive contents. The following case illustrates the problem. X might be a reliable indicator of more than one thing; in certain abnormal circumstances such as in failing light, X might reliably indicate toadstools, whilst otherwise being a reliable indicator of mushrooms. Clearly, representing toadstools as mushrooms is a mistake, and it is not the case that X represents mushrooms or toadstools-in-certain-circumstances, even though X does reliably indicate this disjunctive content. An adequate theory of representation must allow for the fact that mushroom-representations can be caused by toadstools in bad light, or any number of other things, and attribute non-disjunctive contents to such representations.

Simple covariation also faces a second main difficulty; that of accounting for the robustness of meaning, since directly identifying content with cause leaves little room to explain how there can be a mental representation of something that is not there to cause it. Meaning is robust because representations can be tokened in the absence of what they represent; they can stand on their own. We can think about mushrooms without there being any nearby. This is not simply the point that we can misrepresent things as there which are not, as is the case in hallucination. Rather, it is the even more obvious point that we can think about something without the thought having been caused by the actual thing being thought about. Not every mushroom-representation which is not caused by a mushroom is a misrepresentation. If you take a toadstool at dusk to be a mushroom, then your mushroom-representation is indeed a misrepresentation. But if on the other hand, your seeing what you know to be a toadstool causes you to think of mushrooms, or if you just think of mushrooms for no particular reason, or because of some conventional mental association, then your mushroom-representation is not a misrepresentation. Rather, it is a robust tokening; a token of a mushroom-representation which was not caused by mushrooms.

Now if one thing reliably indicates another, then it bears natural information about another thing. Smoke is a reliable indicator of fire, which is just to say that it bears natural information about fire. So trying to give a theory of mental representation in terms of simple covariation is just an attempt to account for mental representation with nothing except natural information. As such, simple covariation is something of a straw-man theory. For we saw earlier when we talked about Dretske's conception of information,

that it could not in itself account for mental meaning. The world is full of causal covariations, but it is clear that mental representations are not simply more of the same. A causal covariation cannot be erroneous. It does not even make sense to say that smoke might misrepresent fire; either smoke covaries with fire, or it does not. In the same way, you could not have a robust tokening of smoke; there's no smoke without fire. It might be possible to construct a theory of mental representation by using the concept of natural information, but simply invoking the concept cannot be sufficient, for an adequate theory needs to account for misrepresentation and robustness.

Fodor's Asymmetric Dependence theory[43] is an attempt to build a causal theory of representation up from natural information. The theory deals with misrepresentation and robustness by giving three conditions which are supposed to be jointly sufficient for mental representation. Condition one is that mental representations covary with what they represent so that, for example, a natural law must link the property of being a mushroom with the property of being a mushroom-representation. Condition two is that some tokens of a mental representation type were actually caused by what they represent, so that mushrooms must appear within the actual history of a person with the ability to token mushroom-representations. The third condition, crucial to the account's ability to deal with misrepresentation and robustness, is that of asymmetric dependency. For a mushroom-representation to represent mushrooms, the efficacy of toadstools in causing mushroom-representations must be asymmetrically dependent upon the efficacy of mushrooms in causing mushroom-representations. For somebody to be able to misrepresent a toadstool as a mushroom, it must be possible for them to accurately

---

[43] (Fodor 1990), chapter 4

represent a mushroom as a mushroom. But the dependency is asymmetric; even if they could not misrepresent a toadstool as a mushroom, they could still represent a mushroom as a mushroom. Toadstools could not cause mushroom-representations if mushrooms did not.

The theory offers a solution to the disjunction problem. Even if mushroom-representations covary with both mushrooms and toadstools at dusk, since the efficacy of toadstools in causing mushroom-representations is asymmetrically dependent upon the efficacy of mushrooms in causing mushroom-representations, and not *vice versa*, then mushrooms and not toadstools are the content of mushroom-representations. The theory also allows for robustness. The idea is that robust tokenings of mental representations are asymmetrically dependent upon non-robust tokenings. A robustly tokened mushroom-representation is about mushrooms because it is asymmetrically dependent upon the efficacy of mushrooms in causing mushroom-representations. In other words, you can think about mushrooms whenever you want because the content of such mental representations is fixed by cases in which mushrooms caused mushroom-representations, and condition two ensures that some mushroom-representation tokens were actually caused by mushrooms.

Naturally, there are many more details to Fodor's theory, but enough has been presented for us to see two important reasons to be dissatisfied with it. The first centres around Fodor's claim that his conditions are sufficient for representation. If anything meets these conditions, then it is a mental representation, and the theory will deliver its content. But not every mental representation need meet the conditions. Representations of

fictional entities will clearly flout condition two, for instance. This does not bother Fodor, since his conditions are not meant to be necessary. Sufficient conditions are good enough for Fodor because, as he puts it,

> ... if we don't know these conditions not to be satisfied, then we would at least be in a position to claim, for example, that "cat" [i.e a cat-representation] *could* mean *cat* for all we know to the contrary ... the prima facie plausibility that "cat" *does* mean *cat* is, after all, pretty substantial.[44]

Fodor is content with this result because if his conditions do cover basic mental representations in which the subject has had direct contact with what their representation is about, then such representations as these could be used as building blocks for the compositional semantics of his Language of Thought hypothesis.

The problem is that there is good reason to suppose that the conditions are not jointly sufficient for representation; it is possible to construct cases in which all three conditions are met by something clearly lacking in semantic value. Fred Adams and Ken Aizawa[45] suggest the following sort of case. In an imagined world, squids produce a certain type of ink. Scientists produce qualitatively identical ink, and can do so only because squids can; squids and scientists are the only ink-producers in this imagined world. If squids could not produce the ink, scientists could not either, and hence the ability of scientists to cause ink is asymmetrically dependent upon the ability of squids. Condition one is met by ink because a natural, causal law links squids and ink. Condition two is met by ink because squids have actually caused ink. And condition three is met by ink because

---

[44] (Fodor 1990), p. 131
[45] (Adams & Aizawa 1994)

the ability of anything apart from squids to produce ink, namely scientists, is asymmetrically dependent upon the ability of squids to produce ink. So since all the conditions are met, if Fodor's theory were correct, then ink ought to represent squids. But it clearly does not. Ink does provide natural information about squids, since it covaries with them, but it does not represent them in the way that a mental squid-representation does. Ink is not capable of misrepresentation, for instance. Fodor's asymmetric dependency condition was supposed to demarcate mental representations from natural information, but cases can be constructed in which it does not, and so Fodor's three conditions cannot be sufficient for representation.

It might be thought that the imagined case is flawed because there could be no reason that the ability of scientists to produce ink should depend upon the ability of squids to do so. But as Adams and Aizawa point out, Fodor does not explain why, for example, the efficacy of toadstools at dusk in producing mushroom-representations is asymmetrically dependent upon the efficacy of mushrooms in producing mushroom-representations. He does not need to because he is only trying to give sufficient conditions for representation, and as is clear from the above quotation, he thinks that it is obvious which way the dependencies would go. But if Fodor does not need to explain the reason for the dependencies he envisions, then Adams and Aizawa do not need to explain why the ability of scientists to produce ink might depend on the ability of squids to do so. However, they could easily do so: perhaps the scientists have so much respect for squids that should squids no longer be able to produce ink, the scientists could not take the task of producing their own ink seriously enough to even try.

A genuine counter-example exists to Fodor's claim that his conditions are sufficient for representation. There is also a second, principled reason to be dissatisfied with the asymmetric dependence theory. Fodor points this out himself:

> Where, in the picture of representation that we've been constructing, does the idea get a foothold that there are *mis*representations; and that they are things to be avoided? (...) There's no obvious reason why the fact that one way of using a symbol is asymmetrically dependent on another implies that we should prefer the second way of using it to the first.[46]

What Fodor's theory lacks is normativity. His theory does not provide a norm against which error is to be measured. It makes the point that robust tokenings and misrepresentations are parasitic upon representations that simply covary with their causes, but it does not say why there is something wrong with misrepresentation. As he says, his theory gives no explanation of why we should prefer accurate representation to misrepresentation, if the only difference between the two is their position within an asymmetric dependency relation.

Fodor has publicly considered adopting a teleological theory in the past,[47] in which normativity would be introduced into a covariation theory by considering the naturally-selected teleological functions of representations. Misrepresentation would then be explained as the tokening of a representation which failed to perform its teleological function. Once a teleological norm is introduced, misrepresenting can be explained as malfunctioning. Fodor abandoned teleological theories because he came to think them

---

[46] (Fodor 1990), p. 128

[47] See (Fodor 1987), chapter 4, and (Fodor 1990), chapter 3. Millikan discusses the "view Fodor would have held had he not thought better of it" in (Millikan 1993), chapter 6 (quotation taken from p. 123).

"rotten at resolving intentional indeterminacy",[48] that is, he found arguments to show that such theories can ascribe indeterminate or inaccurate contents to mental representations. We will take issue with these and other arguments to the effect that teleosemantics cannot make the right content ascriptions in Section X. In that section, it will be argued that Fodor's misgivings about teleology are unjustified. If this is right, then there is no good reason to refrain from introducing normativity into a theory of representation via teleological functions.

Fodor certainly wants normativity in his account, and proposes that it might be introduced "the other way 'round",[49] by starting with the content ascriptions made by the asymmetric dependence theory, and assigning functions to mental representations on the basis of those contents. So for example, a representation which the asymmetric dependence theory says has mushrooms as its content, is assigned the function of representing mushrooms whenever mushrooms are present. If a toadstool causes the representation, then it has malfunctioned. Perhaps this could be made to work, but unless we are moved by Fodor's worries about teleological theories, there is no reason to consider it. Asymmetric dependence theory does not say what error is. Rather, it tells us how to recognise it by giving an account of the structure of error. Teleological theories, on the other hand, are inherently normative; the normativity which explains what error is does not have to be artificially introduced into such theories in the way that it has to be introduced into Fodor's theory.

---

[48] (Fodor 1990), p. 129
[49] ibid., p. 129

The appeal of evolutionary, teleological considerations to the covariance theorist is that they provide an account of the circumstances which determine content for intentional states. The idea is that certain representations have been selected to be tokened in certain circumstances, and so misrepresentations can be explained as tokens produced in the wrong circumstances. Representations would have the teleological function of covarying with whatever it is that they were selected to covary with. The main proponent of this sort of theory is Dretske,[50] who explains the circumstances in which a representation is meant to be produced as those circumstances in which the representation performs the biological function that it was naturally selected for. Intentional states can represent their contents, that being their function, without necessarily indicating the presence of the state's normal cause. Hence, Dretske's theory can cater for the robustness of meaning, as can any teleological theory.

Dretske's theory can also alleviate the disjunction problem, although not as straightforwardly as might be thought. The problem is that teleological functions can be subject to a certain indeterminacy, a point which can be brought out by considering Dretske's example about a type of marine bacteria in the northern hemisphere which have internal magnets called magnetosomes.[51] The magnetosomes guide the bacteria towards the geomagnetic north, which has the effect of leading them away from oxygen-rich water in which they cannot live. It might be supposed that the teleological function of the magnetosome is to guide the bacteria away from oxygen-rich water, since performing this function is what has had survival-value for the bacteria. Nevertheless, it is just as much the

---

[50] (Dretske 1986)
[51] ibid., pp. 164-168

magnetosome's function to guide the bacteria to the geomagnetic north, for within the bacteria's natural environment, the performance of this function has always had survival value. If such indeterminacies can exist for teleological functions, then Dretske needs to give an account of how to distinguish the natural teleological function of mechanisms like the magnetosome. Otherwise, his account would not be able to attribute determinate teleological functions to mental representations, and hence would not be able to ascribe non-disjunctive contents to mental representations, since on his account, function determines content.

Dretske finds that the bacteria are not capable of genuine misrepresentation. This is because he finds it implausible to suppose that if a bar magnet is used to alter the local magnetic field in which the magnetosomes operate so as to make the bacteria swim into oxygen-rich water, then the bacteria are misrepresenting the direction of oxygen-free water. Rather, the magnetosome is simply covarying with the direction of the local magnetic north, and so the question of what it represents cannot be determinately resolved. But with an organism that can receive information in more than one way, for example, a gazelle which can detect a predator by either sight or sound, Dretske thinks the situation is importantly different. It might be thought that there would still be an indeterminacy as to whether the gazelle's representation represents a predator, or just a certain sight or sound, just as there was indeterminacy over whether the bacteria represented the absence of oxygen-rich water or the geomagnetic north. But the gazelle's representation cannot be carrying information about the sound of a predator, for instance,

because the same representation can be tokened without the sound occurring. That is, the same representation can be caused by just the sight of the predator.

But this does not rule out the gazelle's representation being of the disjunction of its two proximal stimuli. To ascribe a determinate content, so that genuine misrepresentation is possible, Dretske thinks that we must consider organisms capable of associative learning. If an animal can learn to associate new proximal stimuli with a given representation, then the content of the representation cannot be the disjunction of the stimuli, since there is no limit to the number of stimuli that could cause the representation to be tokened. In the case of the gazelle, the content of the gazelle's representation cannot be a disjunction of the proximal stimuli which would cause the representation, for there is no limit on the number of possible disjuncts. All that is common to the various stimuli that can cause the representation is the presence of a predator, and so a determinate content can be attributed to the representation on the basis of its having the function of carrying information about the presence of predators. Misrepresentation occurs whenever this function is not performed, that is, whenever the representation is produced in the absence of predators.

The real problem for a teleological covariation theory is not the disjunction problem, but the fact that we have every reason to believe that misrepresentations are not always just malfunctions, and that certain mistakes are, from the perspective of evolution, sometimes meant to happen.[52] Nature errs on the side of caution. Gazelles often bolt when no predators are nearby, and people walking through dangerous areas at night, might react to shadows and noises that would not normally seem threatening. For survival purposes,

[52] See in particular (Millikan 1993), chapter 4.

such jittery tendencies are probably for the best, and generally do not make for more danger. And if certain types of misrepresentation have been naturally selected, teleology will not help in the way Dretske proposes to use it. But teleology still might hold the key to the problem. Natural selection is concerned with results, not means. It is the effect on the survival of squids that explains their ink; it is not that they are meant to produce ink, but rather that doing so has had the right effects. Likewise, it is not that we are meant to produce accurate representations, but rather that the use of accurate representations has effected beneficial results. We need an account which takes account of representation consumption.

VIII - Teleosemantics and Functionalism

Covariation theories define mental representations in terms of their production in response to a state of the world, thus identifying cause with content. This leads to problems when the cause is maverick. Production theories see representation as akin to mirroring the world, but have trouble explaining what counts as an accurate reflection. Systematicity is not enough, as is shown by the disjunction problem: in certain circumstances, James might always represent springboks as gazelles, but he is still misrepresenting gazelles, and not correctly representing springboks in those circumstances. A criterion as to what the contents of representations are supposed to be is needed. A natural suggestion is to invoke teleology, as Dretske does. Adding teleology to covariation allowed Dretske to link representations to those causes which they have been naturally selected to be a response to. But this is a backwards teleology; it is not that certain representations are supposed to be produced in response to certain causes, but rather that consuming those representations in certain circumstances has promoted survival. Natural selection is concerned with results, not means. Squid ink is not meant to be produced in response to danger, rather it is that the result of secreting ink in dangerous situations has been to promote the survival of squids. And so in Millikan's account of mental representation, attention is paid to the evolutionary significance of the consumption of representations, namely survival.[53]

For Millikan, a representation is not at all like a natural sign produced in response to the state of the world. Rather it is a functionally defined entity. It is something which, as

---

[53] (Millikan 1984, 1991, 1993)

used by the system to which it belongs, has an historically determined proper function. A representation performs its proper function when its use achieves the results which have been historically responsible for the replication of that type of representation within the system. So the proper functions of representations are determined by the history of the system that consumes the representations. On an account like Dretske's, it is the teleological function of a representation to indicate a certain state of affairs in the world, since representations are bearers of natural information about the world. But on Millikan's account, it is rather that the representation's function is that of representing certain worldly conditions to the organism's representation consuming device. A representation does not represent in virtue of bearing natural information, but rather in virtue of its consumption having had survival enhancing tendencies.

Millikan talks about the teleological functions of representation production, representation consumption and representations themselves. An organism's representation producing device, or producer for short, has the naturally selected function of producing representations for the organism's representation consuming device, or consumer for short, that will, as consumed, correspond to worldly conditions in such a way as to explain why the consumption of such representations has had survival value. It is the performance of the consumer's teleological function which contributes directly to survival, so fixing content, with the producer's function being that of producing the right sort of representations to do this, and the function of the representation itself being that of effecting the consumer in such a way that it can perform its proper function.

The identification of content with a functional property in Millikan's account, pays heed to Hilary Putnam's insight that,

> ... thinking beings are compositionally plastic ... there is no one physical state or event (i.e., no necessary and sufficient condition expressible by a finite formula in the language of first-order fundamental physics) for being even a physically possible (let alone 'logically possible' or 'metaphysically possible') occurrence of a thought with a given propositional content ... The thing we want insight into is the nature of human (and animal) functional organization, not the nature of a mysterious 'substance', on the one hand, or merely additional physiological information on the other.[54]

The mind-body problem has always been closely connected to debates about mental representation and intentionality within analytic philosophy. Whether as a threat to behaviourism, physicalism or naturalism, intentionality has always presented a stumbling block to reductive accounts of mind, since intrinsic intentionality is exclusive to mental things, and has seemed resistant to non-mental description. Nothing physical shows anything like it, as Brentano said. Once the project of analysing intentionality was undertaken, then the resulting account of intentionality was bound to explain the concept in non-mental terms. For if the analysis mentioned anything mental, then it would doubtless mention an intentional notion and thus be circular. Following widespread commitment to naturalism, an account to fit intentionality into the naturalist's ontology is what has specifically been required. Millikan's theory meets this requirement. By defining representation in terms of historically defined functions, she finds a solution to the problems of misrepresentation, and buys into the most widely held solution to the mind-body problem. Functionalism is compatible with naturalism, as it is supposed that

---

[54] (Putnam 1987), p. 14

functionally defined mental states are physically realised, yet as Putnam says above, functionalists need not specify what the physical realisers of mental states are. Thus functionalism avoids the accusations of chauvinism which are levied at identity theories, on account of their identifying mental states with certain human brain-sets, and so denying that a physically abnormal human, or a different type of organism, could have mental states.[55]

There is a natural way of viewing the teleological theory as an extension of traditional functionalism which has been proposed by David Papineau,[56] who defends a different consumer-based teleological theory from Millikan's. Functionalism makes mental states causal intermediaries between perception and behaviour. Consequently, it deals with the causal but not the representational nature of the mental, for a relation to a distant object could not be part of the internal causal role of a mental state. The input to the functions of mental states is the data provided by the sense organs, and the output is behaviour. Now from this picture, it looks as if representation could be built into the account by widening the inputs and outputs so that they include the world. This widened functionalism would take the content of a belief to be its worldly cause, rather than just a certain sensory input. But this amounts to a covariation theory, and so leads to difficulties with misrepresentation, such as the problem of disjunctive contents. This problem is then solved by saying that the real content of a belief is the worldly condition in which the belief's consumption has provided the evolutionary benefits which explain the retention of the belief-type within the system.

---

[55] (Block 1978)
[56] (Papineau 1993), pp. 56-59

However, the idea that the traditional functionalism cannot accommodate the representational nature of mental states, is by no means uncontroversial. The claim of Functional Role Semantics[57] is that the contents of propositional attitudes can be defined in terms of their causal roles, since for a mental state to have a certain content just is for it to have a certain functional role. This position is closely related to the use theory of linguistic meaning advocated by Sellars, which was mentioned earlier. On that theory, the meaning of a linguistic expression is defined in terms of the use it is put to. Functional Role Semantics applies this insight to the task of providing a theory of content, and defines the content of a representational mental state in terms of its use in the overall system of a subject's thought. So it is the theory's contention that a mental representation such as a belief about gazelles has the content that it does in virtue of the functional role it plays, that role being defined in terms of the representation's causal relationships with its subject's sensory inputs, behavioural outputs, and other mental representations.

The standard objection made against functional role semantics is that it cannot account for intersubjective synonymy.[58] For it is implausible that one person's mental representation might have exactly the same functional role as another person's mental representation. And if that is so, then since Functional Role Semantics determines content by functional role, two mental representations which we might ordinarily suppose to be tokens of the same type would be ascribed different contents. For example, suppose that the presence of a gazelle causes both James and Jayne to believe that there is a gazelle nearby. James' belief leads him to photograph the gazelle, and Jayne's belief leads her to

---

[57] (Block 1986, 1987)
[58] (Fodor & LePore 1991) and (Davies 1995)

shoot at it. So James and Jayne's beliefs have different functional roles, since one belief combined with a desire to photograph a gazelle, and led to James' action, and the other belief combined with a desire to shoot a gazelle, and led to Jayne's entirely different action. So if content is functional role, then the two beliefs must have different contents, even though both beliefs are about the presence of the same gazelle. This sort of approach leads to holism about mental content, such that the content of any one of a subject's mental representations would depend upon the contents of the others.

Ned Block, one of the main proponents of Functional Role Semantics, considers this holism acceptable on the grounds that it would only apply to narrow content. The idea of narrow content can be understood by reference to Putnam's Twin Earth thought experiment,[59] which, in outline, is as follows: two physically identical subjects live in environments physically indistinguishable except for the fact that the substance playing the role of water in the one environment is H_2O, and the substance playing the role in the other environment is XYZ. The moral Putnam draws from this imagined scenario is that the contents of the two subjects' thoughts about water would be different because of their different environments; a physically identical brain state would realise a representation of H_2O in the one subject, and a representation of XYZ in the other. In other words, the wide contents of the representations had by these two physically identical subjects would be different. However, the narrow content of their thoughts about water would be the same, since narrow content is defined as that content which does not depend upon a thinker's environment, but only upon the thinker's internal, functional organisation, which would be identical in the case of physical duplicates.

---

[59] (Putnam 1975)

Block's position is that adopting Functional Role Semantics only jeopardises the intersubjective synonymy of narrow content, and that this is not a problem. The idea is that the narrow content of a representation is extremely fine-grained, for it is determined by the representation's functional role, a many-faceted role involving the subject's other representations. In the above example, James and Jayne's beliefs would have different narrow contents. But this does not conflict with the fact that both beliefs are naturally described as having a content about a gazelle being present. This is because the difference between the roles played by the two beliefs in the overall cognitive systems of James and Jayne would be so minuscule as not to be picked up by public language content ascriptions. Public language content ascriptions like "about a gazelle being present" apply to two representations with different narrow contents by missing the fine-grained semantic distinctions between them.

Block holds a "two-factor" version of Functional Role Semantics. The first factor is the account of narrow content in terms of functional role. But a second factor is needed to explain synonymy between public language ascriptions of content. There must be a less fine-grained notion of content according to which James and Jayne's beliefs do indeed have the same content. This second factor must capture what Block calls "the referential and social dimensions of meaning",[60] and would be provided by a causal theory of reference or a theory of truth-conditions. Block does not commit himself to any particular theory of reference or truth-conditions, but he does mention "indicator semantics",[61] which we have been calling covariation, as one possible reductionist truth-condition

---

[60] (Block 1986), p. 101
[61] ibid., p. 105

theory. And there seems no reason why teleosemantics should not be another candidate. The point is that, in order to account for wide content and intersubjective synonymy, Functional Role Semantics needs to be combined either with a causal theory of reference, or with another naturalistic theory of mental content like the ones we have been discussing. Two-factor Functional Role Semantics is not a direct competitor of the teleosemantic approach, for it claims only that narrow content, not mental content *per se*, can be defined in terms of functional role. Nevertheless, it might be thought that accounting for narrow content in terms of functional role is all that is needed to account for intentionality. An argument against this proposal will be presented in Section XII.

The need for more than narrow functional role is illustrated when Block mentions Gilbert Harman's one-factor version of Functional Role Semantics.[62] In order to account for what Block calls the referential dimension of meaning, Harman adopted a widened functionalism with inputs and outputs that included worldly objects, resulting in just the same widened functionalism that we mentioned earlier, in Papineau's story about the extension of traditional functionalism. But the problem with this sort of widened functionalism, as we have already seen, is that it encounters difficulties in accounting for misrepresentation. If the functional role of a gazelle representation is partially determined by the gazelle that causes it, since the gazelle is part of the input to the representation, then if this functional role determines the content of the representation, there is little room to explain how the representation might be mistokened, as it might be in the presence of a springbok at dusk, for instance. For if the representation is caused by a springbok, then it would have a different functional role to that of a representation caused by a gazelle, and

---

[62] (Harman 1982)

so would have a different content, and not be a gazelle-representation. But gazelle-representations clearly can be caused by springboks. What is needed is normativity to provide a standard against which error can be measured, and this is just what a teleological theory can deliver.

IX - Millikan's Theory

The notion of proper function is central to teleosemantics. A representation performs its proper function when its use achieves the sort of survival-enhancing results that it was naturally selected to produce, hence the proper function of a representation is determined by the history of the system consuming the representation. The type of situation in which a representation performs its proper function for the consumer determines the normal explanation for that representation, as a normal explanation gives the historical reason for the retention of a representation-type. For example, situations in which predators are poised to attack determine the normal explanation of gazelles representing certain sounds behind them as produced by predators, and thus bolting. The function has been retained because it promoted survival in such cases. Millikan calls those cases in which a function is properly performed, its normal condition. So the normal condition for a gazelle's predator-representation is the presence of a predator.

Now when gazelles were first mentioned in connection with Dretske, it was suggested that bolting usually occurs when no predator is around, and so it would seem that the occurrence of "normal conditions", as just defined, might actually be quite abnormal. But Millikan does not mean statistically normal. Rather, she means normal for the function, in accordance with the normal explanation for that function. If representing suspicious sounds as caused by predators occasionally saves gazelles, that can be enough for natural selection to preserve the function, if the false alarms have no particularly detrimental effects from an evolutionary point of view.

With the terminological pieces now in place, a formulation can be given of Millikan's account of mental content. What follows is certainly not Millikan's formulation; a large proportion of her book *Language, Thought, and Other Biological Categories* is devoted to giving that. But the formulation that will be given is enough to meet the requirements of this essay, for our concerns here are far narrower than Millikan's, who puts her theory to a number of different uses. Much of the discriminatory material which she makes use of will be ignored. But in straying from the letter of her account for practical reasons, we stay entirely within the spirit, for Millikan insists that,

> What is needed is not to discover what mental representations *really are* but to lay down some terms that cut between interestingly different possible phenomena so that we can discuss their relations.[63]

Her theory is not supposed to be neat. It is certainly not supposed to give necessary and sufficient conditions for representation, for Millikan is violently opposed to the project of giving conceptual analyses.[64] Rather, her theory is just supposed to promote our understanding of mental representation. However, for purposes of clarification, there can be no harm in our formulating a necessary condition for representation on the basis of her theory.

With these qualifications in mind, Millikan's theory can be formulated:

X represents Y just if Y is part of the normal condition for the consumption of X to perform its proper function in accordance with its normal explanation.

---

[63] (Millikan 1993), p. 104
[64] See in particular (Millikan 1984), chapter 9.

So, for example, a gazelle's predator-representation X represents predators just if predators were present in those historical situations (the *normal condition* for the representation) in which X helped gazelles to survive (performed its *proper function*), thus explaining why the ability to produce that representation type has been retained in the species (the *normal explanation* for the representation).

The immediate advantages of this account are threefold. Firstly, none of the terms used are intentional notions, and are all naturalistic. Secondly, content is determined by past situations which provide the historical explanation for the retention of a representation type. Those situations provide the standard against which misrepresentation is to be measured, and since a representation might have been retained for performing a function which is useful only when very rare circumstances prevail, we can understand why certain representation types might be highly prone to error. Moreover, the account can explain how a function could bring success, enhance survival, and yet still be a misrepresentation. This would happen whenever the function performed by the representation was not part of its normal explanation. Thirdly, by focusing on representation consumption instead of production, representations are no longer depicted as bearers of natural information about their causes, a picture which led to inextricable tangles with the possibility of error. Rather, it is being used by a system in accordance with a normal explanation which makes something a representation; it is the consumer that provides the rules of interpretation by which representations are mapped onto the world. This also serves to demarcate representations from natural side effects which are systematically produced in response to the environment, such as sun-tans. Simple

covariation offered no such demarcation, and so made all reliable natural indicators into representations.

X - Indeterminate Content, Reduced Content, and Other Problems

We shall now proceed to defend Millikan's teleosemantic theory against a number of objections which might be made against it. Doing so will clarify the theory, and bring out some of its finer details for the first time. A good place to start is with a simple charge of circularity, which is as follows. Teleosemantics manages to account for misrepresentation by introducing an evolutionary norm against which error is to be measured, and so defines representation in terms of what it is supposed to do, from an evolutionary perspective. But defining a representation in terms of its proper function, that is, in terms of what it is supposed to do, seems to make the account circular. For surely, it might be argued, there must be something wrong with defining predator-representations as those representations whose function it is to represent predators. But this objection is simply based upon a misunderstanding. Circularity is avoided because the definition of a representation does not just refer to proper function, but also to normal conditions for the representation and the normal explanation for its consumption. Normal conditions and explanations are historical, and so circularity is avoided by referring to actual past events. A predator-representation is something with a function that aided survival in historical situations involving predators.

Another immediate objection to teleosemantics concerns the notion of normal conditions. Since many factors must combine to produce normal conditions for a representation, then simply specifying that some state of affairs which is part of the normal condition fixes content cannot be enough; we need to know which part of the normal

condition it is. For though the presence of a predator is necessary for selectional reinforcement of the gazelle's representation, a lack of mine-fields in the direction of bolt would be equally necessary. Escaping the predator will not promote the gazelle's survival if it is promptly blown to smithereens. So why should the normal condition fix the content of the representation as "predator behind" rather than "no mine-field ahead"?[65]

Millikan's answer is that normal conditions are an idealisation, in that only conditions basic to the "most proximal" normal explanation of the representation's function are considered. It is the factor directly relevant to survival in the normal case which fixes content. Lack of mine-fields is just an omnipresent background condition for survival, which is not specific to the normal condition for the predator representation. Awareness that there is no mine-field ahead cannot explain why predator-representations have helped gazelles to survive in past situations, as if the gazelle only knows that there is no danger in front, it might as well stay put or run towards the predator. It is true that in the normal case there cannot be a mine-field ahead, and it is also true that there must be oxygen in the air, but it was the presence of a predator which defined all of the historical situations where that representation helped the gazelle to survive. Many other representation types need the presence of oxygen and the absence of mine-fields in their normal condition, but the presence of predators in the normal condition is a specific requirement of predator-representations.

It might be thought that this answer would be subject to a similar circularity objection to the one made before, because the injunction that only the "most proximal" explanation should be considered answers nothing. It simply prompts the question, "most

[65] This sort of example is discussed by (Cummins 1989), pp. 77-80.

proximal to what?" to which the only answer could be, "most proximal to what the representation is about." But the answer to this accusation of circularity is just like the one before: content is not only explained in terms of proper function, for the account also refers to actual historical situations. The most proximal explanation mentions the aspects relevant to survival in past situations involving the representation type.

Perhaps the best known objection to teleosemantics is the problem of indeterminate content. This problem was raised by Fodor,[66] who used an example about the fly-detecting mechanisms of frogs; but nothing will be lost by our sticking with gazelles. The problem is that Millikan's account is perfectly compatible with the gazelle's representation having the content "danger", rather than "predator", for if the less specific representation of danger still leads to the gazelle bolting, then it would promote the survival of the gazelle just as well as a representation of a predator would. Both danger and a predator are, after all, equally present within the normal condition for the representation. So it would seem that teleosemantics makes content indeterminate, for since both representing danger and representing a predator would help the gazelle to survive, then either factor from the normal condition might enter into the normal explanation of the representation. There is nothing to choose between the two from the standpoint of survival, and so such considerations seem not to be able to determinately fix content.

This problem can be resolved by remembering that we must always consider the most specific normal explanation for the representation in order to narrow down the possible contents. "Danger" is not enough to define the situations in which predator-

---

[66] (Fodor 1990), chapter 3

representations have survival value, because in many dangerous situations the gazelle would survive by staying still rather than running. But, the objection continues, why not say that the gazelle represents "danger behind that will chase me"; this seems as specific to the normal condition cases as "predator behind". But then, maybe the gazelle knows that predators will not run through water, and so heads for the nearby stream. In that case, the alternative content would need to be, "danger behind that will chase me, but not through water". This could continue for just about everything that the gazelle knows about predators, and eventually the content would become a very complex description of a predator, as experienced over the evolutionary history of the gazelle. The gazelle has a predator-representation, however you describe it.

An objection might be made based on the distinction between types and tokens of representations, as follows.[67] The causal history which fixes the contents of representations must be about types, since content is fixed by *normal* historical situations. But surely this would mean that if the theory were true, then representations would be about types rather than individuals; the teleological theory would make the frightened gazelle's belief into a belief about predators in general, rather than about the particular predator which is behind it. To answer this objection, the notion of normal explanation needs to be further explicated. The normal explanation of a particular representation is that when tokens of that type have been consumed in the past, this has tended to promote the survival of the consumer within certain situations which have fixed the content of the representation. The normal explanation of a particular representation explains why the ability to form a certain type of representation, of which that particular representation is a

[67] This objection was made by J.J. Valberg.

token, has been naturally selected. Now in a particular situation involving a predator, the gazelle forms a belief of a certain type because forming such beliefs had survival value over the gazelle's evolutionary history. But it is not that the belief token is about the type "predator"; rather it is a token of the type of belief that is about predators. The ability that the gazelle has to form the type of belief that is about predators can be thought of as like the ability to form a template for belief that needs input from the current environment in order to generate a belief token.[68] The content of the belief is already fixed by type, ready for a particular belief to be tokened when a particular predator triggers it.

The next objection to be considered is Christopher Peacocke's problem of reduced content.[69] According to Peacocke, teleosemantics must restrict the content of representations to those objects which have causally interacted with the evolutionary history of the representation, for the contents of types of mental state whose consequences have had no causal impact upon the thinker cannot have been naturally selected. Yet many mental states have contents which fall outside of this restriction. For example, beliefs about universal quantification are often meant to range over objects that are causally isolated from the believer, as in the generalisation, "All Londoners are under eight feet tall", which whether true or false, is clearly about all Londoners, and not just those that the believer shares a causal history with. So adopting Millikan's theory would mean jeopardising the meaning of logical terms.

The first thing to notice about this objection is that it applies equally to any theory of representation with a causal component. For instance, covariation would take the

---

[68] This idea was suggested by David Papineau.
[69] (Peacocke 1992), chapter 5

content of a belief about "all Londoners" as covarying with its cause, namely all of the Londoners that the believer has a causal link to. But this will not be every single Londoner; covariation reduces content as well. Since we are finite creatures, it seems that if a theory of representation is not to reduce content, then it cannot be naturalistic. We have already found a number of reasons motivating the production of a naturalistic theory of mental content, and a more thorough case in favour of a naturalistic conception of mind will be built up in the following sections of this essay.

In any case, there is a more clear-cut answer to this objection, which is that the teleological theory treats intentional states as compositional.[70] The content of any given belief has a number of components with their own proper functions. The proper function of each of these components is to systematically contribute in a specific way to the proper functioning of the intentional state as a whole. The meanings of the components are fixed by their contributions to the content of intentional states which have been naturally selected for their contribution to the survival of the organism. Once these meanings and functions are fixed by the participation of the component in a state which has had its content fixed by evolutionary history, the components can be recombined in novel ways in line with the naturally selected reasoning processes of the organism, and thus new contents can be fixed for intentional states which are not causally relevant to survival. So, for the sake of example, the meaning of universal quantification might have been fixed by historical situations in which people had the desire to find all of their offspring before fleeing from an invading army. When the concept of universal quantification was later used as part of the content of a belief, the belief could range over objects which had not had any

[70] The following answer derives from (Papineau 1996).

causal impact on the believer, or even over objects that could not have a causal impact on the person, because part of its meaning was fixed by a situation in which there was a direct causal link between a mental state about "the offspring of X" and every member of the set of the offspring of X.

An oft-cited source of reservation about appealing to natural selection in a reduction of representation is that although there are clearly normal explanations for beliefs and desires which lead to eating and drinking, there seems to be no evolutionary significance to the consumption of complex representations concerned with things like wallpaper and widgets.[71] Perhaps we have been talking about gazelles representing predators not such much out of convenience, but rather out of necessity. Perhaps such examples are used in the teleosemantic literature not so much because of the fact that in these cases the link between the proper function of a representation and survival is clear, but rather because it is only in such cases that there is a link. Once we start talking about everyday human representations, the link becomes tenuous to say the least. For surely it is implausible to suppose that any significant proportion of the mental representations which western people at the end of twentieth century have could be linked to survival. It is hard to imagine how thoughts about philosophy, for instance, could have any survival value. And yet normal human representations are the concern of philosophers of mind, not gazelle representations. It seems that the teleological theory could at best cover a small-subset of human representations, such as desires for water and the like.

This is not the case, since proper functions need not be innate. Humans have evolved to learn from experience. We learn according to principles installed by natural

---

[71] For example (Crane 1995), pp. 185-190.

selection, and there is a normal explanation for the proper functioning of these principles, since they have promoted survival.[72] Normal conditions for everyday human representations are often just those in which the representation will perform the function for which it was learned and retained by the system. The normal explanation for a representation need only have an indirect link to evolution and survival. The link to survival is via our naturally selected systems of reason and thought. Achieving our aims, satisfying our desires, and above all, getting things right, is psychologically rewarding. This is explained by our selectional history; wanting to get things right and satisfy our desires has been hugely successful for humans. When we represent wallpaper properly, the representation performs the function for which it was learned and retained by our cognitive system. Any link to survival comes from the system, as there is a normal explanation for the proper functioning of the principles which it employs. Getting things right aids survival, and having a representation of wallpaper when wallpaper is around is one way of getting things right. Of course, somebody might find it psychologically rewarding to kill themselves, but this is no counter-example. The principles the person employs in committing suicide were naturally selected because of the contribution to survival that their consumption has historically made. The fact that you can shoot yourself does not mean that guns cannot help you to survive.

Millikan's theory can deal with human and animal representations. Moreover, the account can also deal with the intentionality of mental states which are not propositional attitudes. It was promised earlier that the teleosemantic theory could accommodate the possibility that Brentano's thesis is true, and that as such, all mental states are intentional,

---

[72] See in particular (Millikan 1993), chapter 4.

even those like sensations and emotions which we would not ordinarily suppose to be. For example, a pain in the foot could be said to be directed upon the subject's foot because the normal explanation of the pain involves situations in which further damage was prevented through awareness of such damage, the pain's proper function being to alert the subject to the fact that their foot has been injured. The normal explanation for an emotion such as a feeling of anxiety might involve, for instance, situations in which people doubted the direction that their lives were going in, and thus took positive action to correct it. The exact story to be told is not so important as the fact that a story can be told, and so we can understand the sort of thing that the intentionality of such mental states might consist in. Owen Flanagan points out that rather than trying to attribute specific biological functions to emotions, it might be easier to, "... tell a story about the emergence of emotions that gives them a role in enhancing fitness. This is compatible with certain emotional displays being less functional now than they were originally."[73]

Perhaps emotions have no proper functions at all; clearly, not everything is selected for, and not every element of our cognition need have a derived proper function. What is important is just that the teleological theory does not rule out the possibility of attributing intentionality to emotions, and so does not rule out the possibility that all mental states might be intentional. Yet the possibility that mental states other than propositional attitudes could exhibit intentionality is effectively ruled out by alternative theories; to take just the example of covariation, it is very hard to imagine what sort of thing an emotion might be said to co-vary with.

---

[73] (Flanagan 1992), p. 364

The breadth of the explanatory power of Millikan's theory is one main reason why her particular version of the teleological theory is being defended. So far, we have talked about the advantages of the teleological theory, but not about the relative advantages of Millikan's version of the theory. To do this, we shall take David Papineau's theory as the paradigm of the alternatives on offer.[74] The problem with Papineau's theory is that it focuses too exclusively on the intentionality of human mental states. The theory deals with beliefs and desires; the truth-condition of a belief is the condition which guarantees that actions generated by that belief will fulfil their biological purpose of satisfying desires. Papineau thinks that it is a mistake to make the theory wider so that it focuses on survival rather than desire satisfaction, because humans do not aim directly for gene bequest. Rather, humans concentrate on short-term goals, that is, we aim to satisfy desires for things like food and warmth.

Papineau explains that our belief-desire psychology differentiates us from ants who, having no beliefs, are unaware of their environments, and from "super-rationalists" who would have no irrational short-term desires, and thus aim for gene bequests in everything they did. Our psychology is an intermediary in that satisfying short-term goals is good for gene bequest from an evolutionary point of view, if not always ideal. The normal purpose of a belief, according to Papineau's theory, is to satisfy desires, but there are also special purposes by which natural selection has overridden our system of satisfying desires whenever those desires are bad for gene bequest. An example of this is belief in invincibility. Such beliefs are false, but they lead people to go to war and so gain the advantages which fighting has for gene bequest. This is a case where a special function

---

[74] For exposition of Papineau's theory, see (Papineau 1987), chapter 4, and (Papineau 1993), chapter 3.

of belief overrides the normal function of satisfying cowardly desires not to get hurt. True beliefs, for Papineau, can guarantee the satisfaction of desires, but this is not what special beliefs are supposed to do, and so he thinks that an account which includes them by lumping all representations together as having the function of promoting survival would be in error. And this, of course, is just what Millikan's theory does.

But the fact that we do not aim directly for gene bequest does not mean that this goal does not lie behind our actions, and so could not be used to define representation. Our folk psychology of beliefs and desires may just be our way of understanding the super-rationalist project of our naturally selected brains. On Millikan's theory, there is a normal explanation for the proper functioning of special beliefs; they lead to the survival of the fittest and so have been naturally selected. Special purposes are a problem for Papineau because he thinks that representations are always meant to be true. But, as we saw earlier with the example of jittery gazelles whose bolting occurs far more often when there is no predator present than when there is one, representations which are highly prone to error have been naturally selected.

Another problem is that if teleology gives us a reduction of intentionality, then it seems implausible to suppose that a different account would be needed for humans than for other animals. On a teleological theory, it seems highly likely that an account of human intentionality would be much the same as an account of ant intentionality. For no matter how more numerous and complex the intentional states of humans are, they are still intentional for the same reason that the instincts of ants are, which is that they have been naturally selected because of their survival value, resulting in the calibration of the

organism to its environment. On a teleological theory, intentionality would be the same for any evolved creature, for it is only the link between the organism and its environment which has been provided by evolution. But belief psychology cannot be applied to ants. In addition, Papineau by his own admission is reinforcing the categories of folk psychology, whilst remaining uncommitted about its ontology, which to be fair, his theory does not commit him to. But it would seem a far better plan to steer clear of the tangential debate about folk psychology, and at the same stroke, avoid a degree of human chauvinism about intentionality.

We have now discussed a fair number of objections which can be made against Millikan's teleological theory of mental representation, as well as having looked at an alternative version of the theory. But before continuing this defence, we might pause to consolidate the position we have reached through the following direct comparison of teleosemantics with the alternative approaches which have been discussed along the way.

Theory of Content	**Misrepresentation?**	**Counter-intuitive?**
Simple Covariation	Disjunction Problem.	Makes error impossible.
Asymmetric Dependence Covariation	Solution.	No explanation for direction of dependency relations.
Teleological Covariation	Cannot account for misrepresentation that is selected for.	Backwards teleology.
Teleosemantics	Solution.	The problem of no content.

Millikan's teleological theory makes accurate content ascriptions. Most notably, it offers a natural solution to the problem of misrepresentation. But as we see from the last box in

the table, it produces counterintuitive results in the form of what has been dubbed the problem of no content. It is to this problem that we now turn.

XI - The Problem of No Content

Imagine a man walking next to a swamp during an electrical storm. As he stops to look at some mushrooms growing on a rotting tree which is lying in the swamp, he is struck dead by the left branch of a dual forked bolt of lightning. The right branch hits the tree, initiating a chemical reaction in the rotten wood which, by a coincidence of cosmic proportions, creates The Swampman.[75] The Swampman is a molecule for molecule double of the man who died at the moment of his creation, but this is entirely coincidental, the only connection between the two events being their common cause. Swampman's accidental status distinguishes this example from Derek Parfit's examples about teletransportation,[76] in which a replica is created intentionally, for Swampman was not meant to be exactly like the man who died. If he were, an attempt at bypassing the problems he raises for teleosemantics might possibly be made by attributing proper functions to him on the basis of his unusual history. But everything about Swampman is entirely fortuitous.

Swampman offers a challenge to Millikan's theory, for he has no evolutionary history. There never were any normal conditions in Swampman's history which might account for his present stock of intentional states in accordance with a normal explanation, and so none of these states can have a proper function. But if representations are just members of proper function categories, it seems that Millikan will have to produce a very ingenious excuse for Swampman. In fact, she produces nothing of the sort. Neither does

---

[75] The example originates from (Davidson 1987), although the paper is not addressed specifically to the teleological theory. Davidson's position is that Swampman does not have thoughts.

[76] (Parfit 1984), part 3. An exposition of the thought-experiment is given at pp. 199-201.

she reject the example as irrelevant speculative fancy. She just flatly denies that Swampman has any intentional states.[77]

Commentators have tended to regard this result as sounding the death knell of teleosemantics.[78] For is it not obvious that if Swampman comes into existence with a physical constitution indistinguishable from that which the unlucky man was in at the moment of his death, then if the man had a belief about mushrooms, Swampman must have as well? For it is surely not possible for a neural configuration to realise an intentional state one moment, and then nothing the next in exactly similar circumstances. So if Swampman is indistinguishable from the recently deceased man, even at the sub-atomic level, then he would just scratch his head, climb out of the swamp, and continue the dead man's life as though nothing had happened. And if this is so, then to deny that Swampman has intentional states might seem nonsensical.

But the denial would make sense as either a rejection of physicalism, or as a definitional fiat. It would make sense as a rejection of physicalism, for a dualist could argue that Swampman is simply a shell; the machinery without the driving spirit. Less extravagant non-physicalist positions would be of little use here, but it is worth making the point that naturalism or physicalism need not be presupposed in order to embrace the teleological theory. The main reason to be a contemporary non-physicalist is the belief that intentional categories will not and need not reduce to physical categories. But although a major motivation for the teleological theory is the need felt to slot intentionality into naturalistic categories, the theory also wants to explain intentionality, denying that it is an

---

[77] (Millikan 1984), p. 93
[78] A typical example is (MacDonald 1989).

irreducible brute fact about the world which resists further description. A non-physicalist might find this latter aim laudable, and accept an historical, functional characterisation of mental states whilst maintaining that physical description alone cannot capture every fact about consciousness, for instance. However, as was said, such a position is not really relevant to whether Swampman has intentional states, and moreover, Millikan clearly has nothing of the sort in mind. So it looks as though definitional fiat alone has withheld intentionality from Swampman; a decision has been made to define intentional states in terms of historical function, which means that an accidental thinker could not represent anything, no matter what evidence to the contrary there might be.

This position seems unacceptable, for surely there would be something about Swampman's present constitution which would decide the matter; underlying phenomena of the type that would be appealed to in deciding whether something was a toadstool or a mushroom. There was surely something real about the man which made him capable of exhibiting intentionality. And not only is Swampman physically the same as the man; Millikan does not even rule out the possibility that Swampman would be in exactly the same state of consciousness as the man was,[79] and if he were, then Swampman's experience would feel exactly the same to Swampman, as the man's experience used to feel to the man.[80] Swampman cannot tell that his seeming intentional state experiences are not intentional, and brain scientists could not tell either. Surely this means that

---

[79] "Though possibly [Swampman] would be and would *have* to be in a state of consciousness exactly like [the man], that being would have no ideas, no beliefs, no intentions, no aspirations, no fears, and no hopes. (His non-intentional states, like being in pain or itching, may of course be another matter.)" (Millikan 1984), p. 93

[80] Assuming that how experience feels to a subject is determined exclusively by non-intentional qualia. These issues will be discussed in Section XIV.

teleosemantics is just a stipulation, and that the whole project is just a clever way around the problem of misrepresentation faced by covariance theorists. Which is just what many philosophers have suspected.

"How could any "should"s or "supposed to"s be applied to the inner arrangements of your newly arrived, randomly created double?"[81] asks Millikan rhetorically. For it is because it makes sense to talk about what representations are supposed to do, that is, whether they are operating correctly or malfunctioning, that she thinks they must be defined by their naturally selected proper functions. But Swampman has no history which could be appealed to in order to define proper functions for any of his cognitive processes. So since there is nothing he is supposed to do, his brain states are not supposed to be about anything. Nevertheless, we might think that they would be in any case. After all, if he is accidentally just like a man, why should he not also accidentally represent the world? And since the possibility is so far fetched, perhaps it would be best, *contra* Millikan, to defend teleosemantics by finding a way to assign proper functions to Swampman. One way to do this would be by relevant population, in a fashion resembling a proposal once made by David Lewis in the context of traditional functionalism.[82] Swampman is physically identical to a certain man, and so on physical grounds he qualifies for membership of a certain population, namely mankind. Mankind has an evolutionary history which has bestowed proper functions upon human brain states. And so we should attribute proper functions to Swampman on the basis of the evolutionary history of the population which his physical constitution assigns him to.

---

[81] ibid., p. 94
[82] (Lewis 1980)

However, this would be to admit that it is Swampman's present state which determines whether or not he is capable of representation, for it would be to say that his not having an evolutionary history is irrespective so long as he is currently just like organisms which do. It would be to admit that history is not vital to representation, but rather a factor common to all known intentional systems which can thus be used to define representation. This proposal is completely against the grain of the teleosemantic theory, which takes representation as functionally defined in terms of evolutionary history. Unless teleosemantics is to be trite, then as Robert Cummins puts it, Millikan needs natural selection to be more than, "the only available technology for producing ... a system with the right molecular structure."[83]

[83] (Cummins 1989), p. 82

XII - The Case against Accidental Thought

It was Millikan herself who first brought up the issue of accidental thought. This is because she views the possibility not as a problem for her theory, but rather as a vivid demonstration of the consequences of naturalising intentionality. The consequences are a rejection of aspects of the concept of mind which grounds intuitions to the effect that Swampman must be capable of exhibiting intentionality. To accept the theory, regarding Swampman as an unfortunate, if minor consequence, would be to miss the point.

Denying Swampman intentional states lacks intuitive appeal because we know that he is physically the same as a man who did have intentional states. Yet being physically the same does not rule out difference *simpliciter*, the obvious example of this being Putnam's Twin Earth thought experiment.[84] According to externalism, physical constitution does not uniquely determine what an intentional state is about, for the environment in which the contents of a thinker's mental states are fixed has a part to play. Swampman has no history of interacting with any environment; whatever physical state realised a content of H_2O in the dead man, and a content of XYZ in his Doppelgänger, need not realise any intentional state at all in Swampman. Or at least, it is not obvious that it would. Externalists believe that naturalistic, historical relations between thinker and world can determine *what* a mental state is about. The question is, can such relations determine *whether* a mental state is about?

---

[84] (Putnam 1975)

To see that they might, consider how much difference such relations will make to Swampman's purportedly intentional mental states. Nearly all of his alleged beliefs[85] about his past are false. For instance, he believes that he had breakfast with his wife that morning, but he has never eaten anything. He does seem to have true beliefs about world history, his height, and Jayne Cortez's birthday. But he has false beliefs about his learning world history, about measuring his height, and about Jayne Cortez being his wife. Moreover, he has no memories whatsoever, but lots of mental states which seem to him to be memories. For memory requires a past event to be causally connected in the right way to the present representation of that event, and this requirement will always be flouted by Swampman.[86]

Now of the true beliefs he does have, clearly none count as knowledge because none are justified. Swampman only believes them by accident; it just happens that the features of his beliefs correspond to features of the world. But since the correspondence is just a coincidence, there are good reasons to think that his beliefs are not even about the things in the world which they seem to be about. For example, Swampman's belief about Jayne Cortez has no causal link to the woman, so although the phenomenology of his mental state might be just like the late Mr. Cortez's would have been in the same situation, Swampman's belief is not about her. There was an historical, causal reason for the correspondence between woman and mental state in Mr. Cortez's case which is why the state was about her, but for Swampman, any correspondence is just luck. His belief is about a Jayne Cortez who does not exist, and so is false or lacking in truth-value. If this

---

[85] For ease of exposition, we shall just talk about Swampman having beliefs, desires, and other intentional states from now on; of course, whether or not he does have such states is exactly what is at issue.
[86] Here, we take the causal theory of memory for granted, cf. (Martin & Deutscher 1966).

seems odd, consider Mr. Cortez's Doppelgänger visiting Earth from Twin-Earth; his beliefs about Jayne would clearly be about his wife, not the woman on Earth. Swampman's belief is about the wife he does not have.

This point could be extended to his other beliefs; Swampman has no true beliefs after all.[87] And his false beliefs are doubly false. His belief about his breakfast represents a situation in which he had breakfast with his wife, which is false because the situation did not obtain in the world, and also because the elements comprising the representation are not about things in the world. His seeming memories are not about the worldly things that they accidentally correspond to either. Neither are his desires, hopes, and aspirations for the future, for these all seem to be about situations which involve things which he has never encountered, nor has any historical relation to whatsoever. Yet by sheer chance, Swampman says things which ordinary speakers would understand. Consider the case of someone from Twin-Earth being asked on Earth what year the Battle of Trafalgar was, and answering 1805 by applying their memory in the normal way. People from Earth would consider the answer correct, even though it was given on the basis of knowledge concerning a battle which occurred on a far distant planet. The person clearly has no beliefs about the Earthly Battle of Trafalgar, but rather has beliefs which would be mistakenly thought to be true beliefs about the Battle of Trafalgar by someone from Earth. With Swampman, the situation is more extreme, for his thoughts are not about anything at all. The mistake made by people hearing Swampman express a belief about the Battle of Trafalgar would be that of thinking his belief was about any battle at all.

---

[87] Beliefs about abstract and non-existent objects will be discussed in Section XIII.

It might reasonably be thought that the success of these arguments to show that Swampman's intentional states cannot be about what they seem to be about, depends upon two positions in the philosophy of language and mind, neither of which have been argued for. That is, a causal theory of denoting, and externalism about meaning have simply been assumed. The former assumption lies behind the argument that since Swampman has no history of causal interaction with the objects that his mental states seem to be about, then his states cannot be about those objects. With a Russellian analysis of denoting with descriptions, however, it seems that this causal argument would become irrelevant. The latter assumption also lies behind the argument that Swampman does not have the right causal history to intend the objects that his mental states seem to be about. For if we assume internalism about meaning, then the fact that the narrow contents of Swampman's mental states are identical to those of the late Mr. Cortez, would be enough to guarantee that Mr. Cortez and Swampman would have mental states which intended the same objects.

In answer to the first part of the objection, even those who subscribe to a descriptionist rather than a referentialist theory of denoting with descriptions need some causal link between words and objects, otherwise language would be left floating free of the world. Even if the logical form of a sentence such as "the King of France is bald" is that of an existentially quantified sentence, there still needs to be a causal, probably demonstrative, process by which the components of the existential quantification receive their meanings. The most likely avenue by which the descriptionist might import causal interaction into their philosophy of language would be to accept a causal theory of names.

However they do it, sooner or later the descriptionist is going to have to start talking about historical causal relations, and these will be lacking in the case of Swampman. Therefore the import of the argument is not affected.

The second part of the objection is that the truth of externalism has been presupposed. But it is rather the case that, in a sense, the argument of this section so far just is a defence of externalism. The possibility of a Swampman shows internalism in a bad light, for the internalist must attribute specific meanings to mental states which have been realised by accident. If the wind were to blow the sand on a beach so that it formed a sentence, we could interpret the sand as meaning something, even though it is clear that it would not mean anything in itself. Arguably, the same is true of Swampman's mental states and utterances; a certain organic constitution cannot be enough in itself to generate intrinsic intentionality. This is why it is implausible to suppose that Functional Role Semantics might account for intrinsic intentionality simply by accounting for narrow content in terms of functional role. The functional roles of Swampman's and Mr. Cortez's brains are identical, and yet since Swampman has never interacted causally with any worldly objects at all, it seems odd to say that his brain states might be *about* any such objects. To say that an ahistorically realised syntactic item's fulfilment of a certain functional role might be enough to guarantee that it be related to a distal object, seems far removed from any ordinary conception of the intentionality of the mental.

Now that these objections have been considered, we can continue with the argument against accidental thought. The externalist arguments which have been presented are supposed to make it clear that many decisions can be made about

Swampman's mental states by considering their causal, historical relations with the world. Yet the intentional status of these states has not yet been seriously questioned. Within the orthodoxy of contemporary philosophy of language and mind, the question of whether Swampman's mental states are intentional never really arises. We take it that though Swampman's belief about his breakfast is false, it is still about a situation where he has breakfast with his wife, a situation which did not obtain in the world. Yet, it might be suggested, we could just as naturally say that it is a mental state which feels like a belief, has a certain phenomenology, but is not about anything at all. We have granted that the belief is not about a state of affairs in the world, or even composed of elements which relate to the world. It does not appear to be a great step to call the state a pure seeming, and deny that it is about anything at all.

The position under consideration has it that although it seems to Swampman that he has intentional states like beliefs, he actually has no intentional states at all. This suits the teleosemantic position, which makes evolutionary history necessary for intentional mental states. Another position compatible with the arguments of this section, and with teleosemantics, would be that Swampman has no thoughts at all; that he is an unconscious zombie. There are difficulties with both positions. The former forces us to drive a wedge between the phenomenology and the intentionality of mental states, since it commits us to the claim that two subjects might have the same phenomenology of conscious experience, whilst one of the subjects has intentional mental states, and the other does not. Moreover, it faces the difficulty of explaining how things can *seem* a certain way to a being with no intentional states, when *seeming* is an intentional notion.

The difficulty with the latter position, which has it that Swampman is an unconscious zombie, is simply that it is so implausible.[88] The fact that it is implausible that a physical duplicate of a man might be a mindless zombie is exactly what motivates the use of the Swampman example as an argument against teleosemantics, since it is assumed that teleosemantics is committed to such a view. However, if it is tenable to hold the position that everything seems to Swampman exactly as it seemed to the man of whom he is a copy, even though Swampman has no intentional mental states, then we can avoid embracing the counterintuitive position that he is a zombie. We shall discuss whether or not the position is tenable in Section XIV; until then, we shall assume that it is.

[88] Millikan is inclined to the view that Swampman must be conscious, and so is Searle, who says that it would be "impossible" for a physical duplicate of a conscious human to be unconscious. See (Searle 1997), p. 147.

XIII - Towards a Naturalistic Conception of Mind

There is a major obstacle to our denying that Swampman has any intentional mental states at all. For to deny that Swampman's seeming belief about his breakfast must be about something, would appear to commit us to an implausible blanket denial of misrepresentations ever being about anything. Only those states which locked onto what they purported to be about would then count as intentional. If we chose to deny Swampman intentional mental states on the basis that his mental states cannot be about the objects which they appear to be about, then to be consistent, it might seem that we would have to say that any mental state which was not about the objects which it appeared to intend would not be intentional. For instance, on this picture, if Jayne had a belief that her husband was dead, then it would be intentional, but if she had a belief that he was on his way home, we would need to say that her mental state did not exhibit intentionality. Indeed, we would have to say that it was not a belief.

This is totally alien to our way of thinking. We naturally assume that the content of Jayne's belief need not be an accurate representation of how things are in order that it be representational. This assumption is entirely justified. But it is not incompatible with denying that Swampman's mental states are intentional. It is not necessary to deny that misrepresentations are ever about anything, since we have an alternative account of misrepresentation in Millikan's theory. In fact, one of the theory's great strengths is its ability to deal with error.

The argument is this. Swampman's mental states could not be about the individuals that they were about in the case of the dead man. Considering his accidental status, it might be thought that it would be better to say that his mental states are not about anything at all. But the argument against this proposal is that mental states could not plausibly be said to lose their intentional status whenever they fail to latch onto the individuals they intend, or we would have to say, for instance, that when the gazelle is wrong about there being a predator behind it, then its mental state is not about predators, which is absurd. Surely, the mental state is about there being a predator behind, even though the gazelle is wrong because that state of affairs does not obtain. But then the proposal was that we use Millikan's theory to explain such clear cases of misrepresentation. Not every seeming intentional state is about something, but those with historically determined proper functions are. It is true that mental states do not always perform their proper functions, and in these cases there is misrepresentation. But when there is no proper function to perform, as is the case with Swampman, then nothing is represented. Misrepresentation occurs when a state does not lock onto to what it is supposed to lock onto, but Swampman's states are not supposed to do anything at all, let alone lock onto something. So if we are to insist that Swampman's mental states do exhibit intentionality, we will need another reason to do so, other than the requirement of being able to account for misrepresentation.

Given that we have a naturalistic alternative to explaining misrepresentation, why else would we want to attribute intentionality to Swampman? His lack of history has ruled out his mental states being about what they look to be about, so why should they be about

anything at all? The reason cannot be outside of Swampman's consciousness. The broad content arguments of the previous section were designed to establish that if Swampman's mental states are about anything, then they are not about the worldly entities that the dead man's mental states were about. If we accept this, then Swampman's present constitution and present causal relationships with the world cannot be exclusively determining what his mental states are about, for exactly the same constitution and present relationships determined something different in the case of the dead man. This suggests that we need to introduce past causal relationships, but in Swampman's case we cannot. The physical world and its history, which in this context is everything outside of his consciousness, suggest that Swampman's mental states are not about anything. At this point of course, you might just insist that a brain state X realises an intentional mental state, and that if Swampman has X, then he has an intentional mental state. But reasons are being given to resist this, namely that we do not need to say this in order to explain misrepresentation, and so the view has no clear motivation. Worse than that, the view is unable to explain what it is that Swampman's mental states are about, if they must be about something or another.

So maybe the reason that we should attribute intentionality to Swampman is to do with what is within his consciousness; it is that Swampman experiences the intentionality of his mental states. Our commitment to this idea must be what makes it seem so counterintuitive to allow the question of whether Swampman's mental states exhibit intentionality to be decided by a definition based on causal and historical matters of fact. For we naturally assume that if a mental state seems to be an intentional state, it must be

about something. We know immediately and infallibly if we have a belief or a desire, and beliefs and desires have contents; they are about something, if not necessarily something worldly. Propositional attitudes are guaranteed to be intentional, for it is part of the concept of a propositional attitude state that it should have a content. Consequently, since we can know when we have a propositional attitude state through introspection, then we can also know that we are thinking about something.

Yet this is exactly the picture Millikan wants to oppose. Thoroughly naturalising intentionality involves rejecting that anything is given to consciousness with certainty. For the whole notion of the transparency of consciousness is a throw-back to the Cartesian image of the subject within a sphere of imminence where everything can be known with certainty, trying to secure knowledge of the sphere of transcendence which is the world. Throughout the twentieth century, philosophers have been trying to shed this image, and root consciousness firmly in the world. And this is exactly what Millikan's theory does so well. According to her theory, representation is a natural relation between mind and object. Any awareness of the intentionality of a representation is nothing more than the inside of that natural relation; the feeling which tells the thinker that an intentional relation is supposed to hold. But it is no guarantee whatsoever that the relation does in fact hold, which is borne out by Swampman, who has all of the feelings associated with intentional states, but not the states themselves. This is not to say that anybody seeking to naturalise intentionality would want to reject that anything is given to first-person reflection. Naturalising intentionality does not threaten personal authority about how we feel. The point is just that how we feel does not guarantee anything about the world. All that we can

be sure of is that we feel as though we have an intentional state, not that we actually have one. To have a content is part of the concept of a belief, but the question of whether a mental state that feels like a belief has a content cannot be settled by introspection.

It perhaps seems odd to blame intuitions about whether a physical replica of a man would have mental states which exhibit intentionality on the Cartesian image of mind. For surely the intuition that physical replicas must be replicas in every way is a product of physicalism, a view at the opposite end of the philosophical spectrum to Cartesian dualism. Yet oddly enough, standard intuitions about the Swampman case do seem to be based upon a hybrid of physicalist and Cartesian intuitions. Our physicalist intuitions amount to the idea that since beliefs are realised by brain states, whenever the right brain state is tokened, then there will be a belief X. Then we assume that Swampman can tell whether or not he has belief X by introspecting, and since beliefs are intentional states, if he does have X, then he can be sure that his thought is about something. It is part of this latter step that is blamed on Descartes; the idea that we immediately and infallibly know whether we have a belief or a desire simply through reflection. According to this idea, the transparency of the mental allows us to know indubitably when we have a belief, and since having a content is part of the concept of a belief, then we can know that the thought has content by introspecting. Modern physicalism and Cartesianism certainly do not mix well. But intuitions about the Swampman case do seem to rest upon this confused concoction.

There is nothing in favour of Swampman having intentional states except confused intuitions based upon an incoherent conception of mind. Swampman should form no basis for a rejection of teleosemantics. Rather, his lack of intentional states should be counted in

favour of a thoroughly naturalistic position which takes human situation within a world as fundamental. The idea that there is nothing more to having an intentional mental state than having the right neural configuration and phenomenology of experience must be rejected, for intentionality is a relationship between a mental state and the selectional history of the subject of the state. A physical make-up realises a phenomenology of experience, which tells the subject what their mental state is supposed to be about. But the question of whether it really is about anything at all is decided by history.

However, it remains unclear how this argument could apply to representations of abstract objects or fictional objects, where considerations about causal interactions cannot be so easily appealed to. For instance, if Swampman has a belief that a square and a circle cannot coincide, then surely it cannot be that his mental state is not about the properties of squares and circles. It seems that nothing which has been said so far applies to this important class of mental states, for it cannot be plausible to say that his mathematical beliefs are not about numbers, for instance. And if this is so, then perhaps the move to deny that Swampman's accidental brain could realise any intentional states was made rather too quickly.

The teleological theory can resolve this problem by reading statements about abstract objects as existentially quantified statements, in accordance with the methods of F.P. Ramsey.[89] Just as the meaning of a term for an unobservable object like an electron gets its meaning from the theory in which it appears, so do mathematical objects get their meaning from mathematical discourse. Our intentional states can be about objects which we accept because we accept existentially quantified statements, for there is a normal

[89] This idea is taken from (Papineau 1993), chapter 6, especially pp. 189-191.

explanation for our reliance upon such mechanisms. Now this normal explanation will be lacking in the case of Swampman, and so his mental states will not be about numbers, squares or circles. Hence, the argument is not invalidated by the possibility of thoughts about abstract objects. If abstract objects really exist, reflection alone cannot guarantee a relation with them. A similar story could be told for beliefs about non-existent objects like Pegasus, whereby intentional states about such objects have derived proper functions from our learning stories in which they are mentioned.

XIV - Intentionality and Consciousness

We now need to consider whether it is tenable to hold that Swampman is conscious even though he has no intentional mental states. For if we are forced to conclude that Swampman is unconscious, then the example would no longer stand as an evocative illustration of intentionality, as naturalistically construed by teleosemantics, but would rather just be a counter-intuitive consequence of the theory, as it is often supposed to be by the theory's proponents and detractors alike.

If Swampman does not have intentional mental states, then it is not obvious that he would nevertheless still be conscious. There is, it can be argued, no reason to assume the phenomenology whilst questioning the intentionality, when it is natural from the point of view of physicalism to assume that a physical replica must have both. After all, there is an intuitively obvious link between the intentionality of a mental state, or what it is about, and the conscious experience of having a mental state, or what it is like. The link is that what a mental state is about determines what it is like for the subject to have that state: if S has a belief with the content X, then it feels to S that he or she has a belief that X. Moreover, a view has recently emerged in contemporary philosophy of mind, of which Michael Tye is a major proponent,[90] to the effect that consciousness depends upon representation, so that one must have intentional mental states in order to be conscious. On this view, then, Swampman must be a zombie.

Alternatively, however, we could take the view that the phenomenology of conscious experience is entirely determined by non-intentional qualia. And there are

---

[90] (Tye 1995), especially chapter 5

reasons to think that such an assumption might well be justified. Representation is a normative concept, which explains the utility of teleological considerations in the provision of a theory of mental representation; teleology provides the norm against which accurate representation is to be measured, allowing misrepresentation to be explained as malfunctioning. The phenomenology of conscious experience, on the other hand, is not obviously normative, and it is natural to suppose that only a certain present state physical constitution is required for its realisation. It seems plausible that although content might depend upon historical functions, phenomenology of experience just depends upon neurophysiology.

The question of the relationship between intentionality and consciousness is a complex and controversial issue. But it is enough for our purposes here, to note that if we are prepared to think of the phenomenology of experience in terms of non-intentional qualia, then we could adopt teleosemantics without denying that Swampman would be conscious. And even if it turns out that such a conception of consciousness is untenable, and that since intentionality and consciousness are so intimately linked, Swampman must be a zombie, this would not render the argument of Section XII irrelevant. For in that section, we found that the assumption that Swampman would have mental states just like the man that he is physically identical to, would not stand up to close inspection. From this, we concluded that it only seems to Swampman that he has beliefs and desires, but that in fact he has no intentional states. But if we held the view that the phenomenal or felt aspects of consciousness are intentional, then the broad content arguments of that section would have given us good reason to think that Swampman is a zombie.

However, we are still not entitled to claim that teleosemantics is even compatible with Swampman being conscious, for we have not yet considered the charge that it is straightforwardly contradictory to hold the position that Swampman has no beliefs even though it seems to him that he does have beliefs, since seeming is an intentional notion.[91] To say that it seems that things are a certain way to Swampman, is just to say that he is representing things as being a certain way. When it seems to Swampman that he has a belief, his mental state must be about the phenomenology of his experience. So surely the argument is entirely self-defeating, so goes the argument, since if things seem a certain way to Swampman, that is, if his experience has a certain phenomenology, then he must have intentional mental states after all. So once more, either Swampman is a zombie, or his physical constitution guarantees that he has the same intentional mental states and phenomenology of experience as did the dead man of whom he is a duplicate.

Seeming is indeed an intentional notion, and so although it could truly be said of the dead man that things seemed a certain way to him, the same cannot be said of Swampman. But we do still need some way of making the point that Swampman's brain will generate just the same state of consciousness in him, as the physically identical brain of his duplicate did. One way of doing this is to start talking about "mock" intentional states; when Swampman has a brain state which would have realised a certain belief in the dead man, then Swampman will have a mock-belief, associated with exactly the same phenomenology as a real belief would be, but lacking any intentionality.

This reply is rather *ad hoc*, but it must be remembered that Swampman is a very special case, stipulated into existence in order to show the consequences of naturalising

---

[91] This objection was made by Timothy Williamson.

intentionality, and to confront non-natural intuitions about mind as starkly as possible. It might seem petty to refuse to credit this fully functioning, rational agent with intentional states, just because of his background. But really, the making of this seemingly petty refusal is the whole point of the exercise. It vividly illustrates that representation depends upon history, and helps us to the see the confusion of the common-sense intuition that present constitution must determine content. Talk about it "seeming that" something is the case for Swampman cannot be taken literally. For consider the case of the man of whom Swampman is a copy having a belief about his breakfast. The man might reflect on how he feels, and have a belief about feeling as though he has a belief about his breakfast. It seems to the man that he has this belief, so this mental state is about the phenomenology of his experience, and hence an intentional state. But Swampman's parallel thought is not. It is not about anything. It is not that it seems that it is about something, but rather that the same physical realiser for Swampman's non-representational thought would realise a mental state in the man of whom he is copy, about it seeming that his thoughts were about something.

XV - The Status of the Theory

> ... some argue that content is a matter of evolutionary history in the sense of being what some structure is selected for. This may be true for some notion of content, but is implausible for representational or informational content. The information we give out when we utter sentences like 'Jones hopes that there is not a tiger behind her' or 'The bus leaves at six' is not information about evolutionary history.[92]

This quotation comes from Frank Jackson, who offers an objection to teleosemantics quite unlike any of those that we have considered so far. Up until now, we have dealt with supposed problems for the internal workings of the theory, the ascriptions of content that it makes, and the consequences of accepting such a theory. We have found that teleosemantics is coherent, makes the right content ascriptions, and has a wide-ranging naturalism as its consequence. Jackson's objection, however, is concerned with the status of the theory. The objection, in outline, is that a philosophical analysis of mental representation should come in the form of a piece of conceptual analysis, and that since teleosemantics cannot claim this status, then it cannot be a theory of the sort of content that interests philosophers of mind. At best, it could only be a theory of some other notion of content: "teleological content".

Conceptual analysis and theoretical reduction are two standard options for the status of a philosophical theory. Familiar examples of each are as follows, where (1) is a conceptual analysis and (2) is a theoretical reduction:

(1) Having a true and justified belief that X is necessary and sufficient for knowing that X.

(2) Lightning is an electrical discharge.

---

[92] (Jackson 1997), p. 273

Conceptual analyses offer necessary and sufficient conditions for the application of concepts, in order to clarify those concepts. (1) is supposed to tell us what knowledge is, by providing a logical equivalent of knowledge which makes no mention of the target concept itself. The correctness of a conceptual analysis is judged by whether the applicability conditions of the analysis are co-extensive with those of the target concept. Intuition provides the yardstick; to see if (1) was true, philosophers asked themselves if they could imagine a case of justified true belief which they would not classify as a case of knowledge, or *vice versa.*

A theoretical reduction like (2) involves the identification of the phenomena picked out by everyday concepts, with phenomena picked out by concepts which play roles in scientific theories. In (2), the reference of "lightning" was fixed by the role played by a certain phenomenon. The phenomenon playing the role was discovered to be electrical discharge, a concept which appears in scientific theories. So the phenomenon picked out by the everyday concept of lightning turned out to be identical to the phenomenon picked out by the scientific concept of electrical discharge. Such reductions are typically tested empirically, rather than by appeal to intuition; trying to imagine a case in which lightning is not an electrical discharge, would not be the way to go about objecting to (2).

In a paper co-written with David Braddon-Mitchell,[93] the argument touched upon in the above quotation from Jackson is made explicit. For Braddon-Mitchell and Jackson, providing conceptual analyses is a central task for the philosopher. Conceptual analyses offer illuminating accounts of what implicitly or explicitly guides someone who understands a concept when they classify something as an instance of that concept, and

---

[93] (Braddon-Mitchell & Jackson, unpublished)

should make sense of the theoretical roles played by the concept under analysis. Braddon-Mitchell and Jackson consider mental content an ideal candidate for conceptual analysis; such an analysis would say what implicitly or explicitly guides us when we identify contents in everyday life, and would do so in such a way as to illuminate the role of content in philosophical theories.

However, it is clear that we do not use concepts connected with evolutionary history, such as those mentioned in the teleological theory, when we identify contents. A conceptual analysis of knowledge can guide attributions of knowledge: for example, we might decide that a person does not know that X, because their belief that X is not justified. Likewise, a conceptual analysis of content ought to tell us what explicitly or implicitly guides attributions of content. But facts about naturally selected proper functions cannot be guiding attributions of content, since people who know nothing of such things have no difficulty in correctly attributing contents such as the belief that X to other people. Braddon-Mitchell and Jackson conclude that teleosemantics is not a piece of conceptual analysis. Teleological theorists do not lament this fact. Papineau thinks that it should have been "clear from the start"[94] that his teleological theory of representation was not meant to be a conceptual analysis of content, and Millikan hints at her general opposition to the project of conceptual analysis when she calls it, "a confused program, a philosophical chimera, a squaring of the circle, the misconceived child of a mistaken view of the nature of language and thought."[95] Both think that the teleological theory is best

[94] (Papineau 1993), p. 93
[95] (Millikan 1993), p. 15. See (Millikan 1984), chapter 9, for her defence of this view.

taken as a theoretical reduction. Moreover, Braddon-Mitchell and Jackson agree with them.

But since the teleological theory is a theoretical reduction, then according to Braddon-Mitchell and Jackson, the content it identifies cannot be what they call "informational" or "representational" content, that is, it cannot be the ordinary, everyday notion of content which interests us. The informational content of a subject's belief is the information about how a subject takes the world to be, as regards that particular belief. An example Braddon-Mitchell and Jackson give is that the informational content of Jones's belief that there is a tiger nearby, is what Jones tries to convey when she says to someone, "there is a tiger nearby". Now it might seem that this intuitive notion of content is exactly what we have always been working with, and exactly what the teleological theory gives a theoretical reduction of. Braddon-Mitchell and Jackson do not argue with this. Rather their point is that theoretical reductions have to live alongside the informational roles of their target concepts. They give the example of the theoretical reduction, "Water is H_2O". When the sentence "There is water nearby" is understood by somebody who knows nothing about chemical structure, then the information which has been conveyed cannot be that H_2O is nearby. Rather it is must be something like the information that, "the clear, potable stuff of our acquaintance which falls from the sky, fills the ocean and is known as 'water' in our language community is nearby".[96]

The theoretical reduction of water to H_2O did not make the informational role of the concept of water redundant, and so, according to Braddon-Mitchell and Jackson, neither can the teleological theory make the informational role of the concept of mental

[96] (Braddon-Mitchell & Jackson, unpublished)

content redundant. At best, the teleological theory can claim to have identified a teleological notion of content, whose ascriptions shadow everyday ascriptions of informational content. But the remarks of teleological theorists are, "evangelical rather than ecumenical"[97]: the teleological theory claims to be *the* correct theory of content, rather than a theory of one notion of content which is of interest because it is defined in terms which figure in evolutionary theory. Teleological considerations are supposed to provide the only possible explanation of the normativity of content, and thus only the teleological theory is supposed to be able to solve the problem of distinguishing accurate representation from misrepresentation. But Braddon-Mitchell and Jackson think that this cannot be so. For we endlessly and effortlessly "solve" the problem of misrepresentation in everyday life. For instance, the problem of disjunctive contents poses no dilemma in our everyday content ascriptions. We do not need to appeal to teleological considerations to tell that Jones has a belief about gazelles, and not a belief about gazelles, springboks at dusk, or anything else that would fool Jones.

We can easily identify contents in everyday life. We do not appeal to the teleological theory to make these identifications, which was only to be expected since the theory is not meant to be a conceptual analysis. However, since the theory is not a conceptual analysis of informational content, it cannot be a theory about the concept of informational content at all. The teleological theory is a theoretical reduction which identifies the phenomenon playing the role of informational content with a phenomenon that has a role defined in terms of evolutionary theory. The account of content it gives concerns the role defined in terms of evolutionary theory, and not the concept of

---

[97] ibid.

informational content. Just as chemical theorems about H_2O are not about the everyday concept of water, so the teleological theory is not about the everyday concept of content. Yet the everyday concept of content, informational content, is what we are interested in, and not some other teleological notion of content. Since any theoretical reduction would be subject to this very same charge of identifying a distinct notion of content from that of informational content, then it is clear that Braddon-Mitchell and Jackson think that any useful account of content would have to be a conceptual analysis. For we can already identify contents accurately; what we need is to know what explicitly or implicitly guides these identifications, and we need to be told in a way which makes sense of the roles played by content in our theories.

It is only because the teleological theory has to be read as a theoretical reduction that it cannot be a theory of informational content. So it might well be thought that the problem applies equally to any of the naturalistic reductions of mental content which we have been considering. Well, this depends on whether the theory in question does need to be read as a theoretical reduction. Although it is generally assumed that naturalistic theories of content are meant to have the status of theoretical reductions, it is not hard to imagine some of them being defended as conceptual analyses. For instance, it could be argued that considerations about covariation implicitly guide attributions of content: we apply a belief about gazelles being present to Jones because we think that she can see the gazelle, and that it will cause a certain representation in her. Maybe it could even be argued that considerations about asymmetric dependency relations guide content attributions as well. But not so in the case of the teleological theory, which cannot be

conceptual analysis on the grounds that people can identify contents without knowing anything about naturally selected functions. So the problem posed by Braddon-Mitchell and Jackson does have particular application to the teleological theory.

The key to Braddon-Mitchell and Jackson's criticism of the teleological theory is their point that theoretical reductions must live alongside the informational roles of their target concepts. In this, they are surely right. But they suppose this point, of itself, to be damaging for the teleological theory, and it simply is not. Braddon-Mitchell and Jackson say that the comments of teleological theorists are "evangelical rather than ecumenical", in that teleosemantics is supposed to provide the only means of making accurate content ascriptions. From this, they conclude that teleosemantics claims to be the theory of one univocal notion of content, and so forgets the fact that contents are routinely identified every day without recourse to teleological considerations. But Braddon-Mitchell and Jackson fail to allow for the fact, which they continually emphasise, that the teleological theory is a theoretical reduction. For theoretical reductions always take the informational roles of their target concepts as a starting point. The evangelical pronouncements of teleological theorists amount to nothing more than the claim that their theory is the only theory capable of making content ascriptions which match those that we are naturally able to make in everyday life. The point is not that teleosemantics offers the only way of accurately identifying content, but rather that it is the only theory which can shadow everyday identifications of content.

There has to be an everyday way of recognising phenomena in order that those phenomena can be theoretically reduced. But once the reduction has been affected, more

discriminatory material is provided to determine which phenomena really do fall under the everyday concept; theoretical identifications can guide and correct the everyday methods used to identify their target concepts. Consider what would happen if an underwater spring full of XYZ was discovered. The people who made the discovery would doubtless believe that they had found an underground spring full of water. But if scientists ran tests on the liquid, and announced that it was XYZ rather than H_2O, then the people who made the discovery might judge that the spring had not turned out to be water after all, that is, they might judge that the liquid in the spring was not actually "the clear, potable stuff of our acquaintance which falls from the sky, fills the ocean and is known as 'water' in our language community", but rather another very similar sort of stuff. People do often defer their classifications of phenomena to theoretical reductions.

Now it is true that once a theoretical reduction of a concept is made, its informational role remains; we do not identify water via its chemical structure in everyday life, and neither do we identify contents via teleological functions. But appeal to informational role is just a way of identifying a certain phenomenon. Once a theoretical reduction is in place, that same phenomenon can be identified in another way, that is, via the theory invoked by the reduction. What is more, when there is a conflict between identifications made by informational role, and identifications made by the theory, then the identifications of the theory would take precedence. Theories of natural science can tell us when our ascriptions of the concept "water" are accurate, and it is the contention of teleological theorists that their theory can tell us when our ascriptions of mental contents

are accurate.[98] Chemical theories do not illuminate the everyday concept of water, and neither does the teleological theory illuminate the everyday concept of informational content. But this is not what they are meant to do; they are not meant to provide conceptual analyses. According to Braddon-Mitchell and Jackson, this means that they are not addressed to the everyday concepts of water and content. Perhaps not; but they are addressed to the phenomena that those concepts are used to pick out. Braddon-Mitchell and Jackson's charge is only important if what is of interest about content to philosophers is the everyday concept of content, and not the phenomenon which it is applied to.

Clearly, there are cases where conceptual analyses are appropriate, and cases where theoretical reductions are appropriate. A conceptual analysis of knowledge might be interesting, but a conceptual analysis of water would not be; we might expect to produce a theoretical reduction of water, but not so in the case of knowledge. With mental content, the case is not so clear-cut. A conceptual analysis of our everyday notion of content might be illuminating, and might be brought to bear upon other theories which make use of the notion of content. But it would make no contribution to the philosophical enterprise of understanding intentionality. The issue of finding out about what explicitly or implicitly guides attributions of content is tangential to the metaphysical issue of explaining what it means for a mental state to be about some state of affairs. But a theoretical reduction of content, on the other hand, can explain what intentionality *is*; by offering a naturalistic account of what it means for a mental representation to have a certain content, such a theory shows us how intentionality fits into the natural world.

---

[98] David Papineau states clearly his belief that the teleological theory might overturn everyday content ascriptions at (Papineau 1993), pp. 93-94.

The answer to Braddon-Mitchell and Jackson's charge, then, is that there is room for both a conceptual analysis and a theoretical reduction of content. They are right that the teleological theory, being a theoretical reduction, does not tell us anything about the everyday concept of content itself, but are wrong to conclude from this that the theory is not concerned with the sort of content that is of concern to us. For the theory tells us about the phenomenon which this ordinary concept of content is applied to, rather than telling us about the concept itself, as a conceptual analysis would do. It is the goal of gaining a fuller understanding of the phenomenon itself, and of reconciling intentional phenomena with the demands of naturalism, which has inspired the production of theories of content like teleosemantics, and so teleosemantics has the right status to deal with the problematic that it is addressed to.

XVI - Conclusion: Trading Intuitions

Teleosemantics is often considered to be an implausible theory. Searle, for one, cannot help thinking that the appeal to natural selection is just a trick for getting around the problem of accounting for misrepresentation. Hence, in a passage where he runs through various theories of content which have been proposed, teleosemantics receives the following précis:

> ... an even less plausible account, is that intentional contents can be individuated by their Darwinian, biological, teleological function. (...) It should be no surprise to anyone that these attempts do not even get off the ground.[99]

The problem of no content is frequently invoked as exemplifying the implausibility of teleosemantics. Hence we find Fodor noting with incredulity that he has,

> ... heard Ruth Millikan claim that it's sufficient to make Swampman's head contentless that he lacks an evolutionary history.[100]

In this final section, then, we shall ask whether teleosemantics really is the sort of theory that common-sense should not allow us to take too seriously, or whether to the contrary, the picture presented by teleosemantics is entirely plausible, once properly taken account of.

Intentionality is a genuinely puzzling phenomenon. As we noted at the outset, the ability of mental states to be *about* distal objects is unparalleled. The challenge to

---

[99] (Searle 1992), p. 50
[100] (Fodor 1994), p. 126, appendix B, note 1

philosophers interested in this phenomenon is to resist positing intentionality as a *sui generis* feature of the world, and instead, to explain how it fits into the natural world. However, most of the theories which attempt to meet this task, appear to offer little or no insight into the phenomenon. Consider covariation, for example, and suppose that representations do covary with their causes. Does this supposition provide a satisfying answer to the question of how a mental state can be intentionally related to a distal object? Not at all; this might well be a fact about intentional states, and a consequence of the existence of intentionality, but it is hard to imagine intentionality *consisting* in this. Of course, this is not to say that the theory is not true. Puzzlement over how mental states can be about things might well be bogus, and covariation plus asymmetrical dependency, for instance, might be all there is to intrinsic intentionality.

But by appealing to teleological considerations, we get the following account of intentionality, which does seem to provide a satisfying explanation of the phenomenon. The general metaphysical insight underlying the teleological theory of representation is that intentionality is the upshot of evolution. Mental states exhibit intentionality in virtue of having been calibrated to fit the world by the random forces of natural selection. The representational contents of the mental states of organisms were selected for, because they contributed towards successful, or rather, survival-enhancing interactions between organisms and their environments. So intentionality can be understood as the relation holding between representations and what they represent, which has been built up by, and is defined in terms of, selectional history. The teleosemantic theory builds upon this insight, by saying that contentful states are historically defined functional states. This

position yields an explanation of intentionality, and at the same stroke, a naturalistic position on the mind-body problem, since the historically defined functional states are understood to be physically realised.

The point to be made is not that the teleosemantic theory's ability to offer a more satisfying explanation of intentionality than its competitors should count as a crucial recommendation for its adoption. The theory's theoretical advantages, and thoroughly naturalistic consequences provide a stronger recommendation. And besides, as we have already noted, any mystery that we feel surrounding the intentionality of the mental might just be a sham. Rather, the point is that since teleosemantics does offer real insight into, and a satisfying resolution of, the problem of intentionality, then the theory should not be dismissed as arbitrary and implausible.

The genesis of the theory has a lot to do with explaining why it is routinely dismissed as implausible. There was a technical problem in providing a theory of representation, which was the problem of accounting for misrepresentation. To solve this problem, normativity had to be accounted for. This led philosophers to appeal to natural selection. Now before the technical problem arose, nobody might have suspected that natural selection had anything to do with intentionality, hence the widespread suspicion that the theory must just be a clever way around the problem of misrepresentation. The situation was not helped by intuitive stumbling blocks like the problem of no content. However, it has been argued here that when evolutionary considerations were introduced to solve the problem of accounting for misrepresentation, they turned out to offer a natural and satisfying explanation of intentionality. It was also shown earlier why the fact

that the theory would not attribute intentional mental states to an accidental duplicate should not be counted against it. So we have a reasonable case from which to conclude that accusations of implausibility levelled at the teleological theory are not justified.

We started out by looking at attempts made to gain further understanding of intentionality. In the process, we came to see the philosophical problems surrounding the concept of intentionality, and the various options that could be taken in the attempt to solve these problems. We found that the best way to proceed, in order to resolve these problems, was to produce a theory of mental content. In considering various theories of content, we found that Millikan's teleosemantic theory offered the only satisfactory solution to the problem of accounting for misrepresentation. We then proceeded to consider a long list of objections which have been made, or could be made, against the theory. In every case, we found either that the theory could straightforwardly deal with the problem posed, or that it clearly had the resources from which to formulate an adequate answer. We then looked at the problem of no content, often considered the most damaging objection to the theory, only to find that the example invoked by the so-called problem, was actually a vivid illustration of one of the theory's central advantages, namely the thoroughly naturalistic conception of mind which it presents us with. Finally, we considered the status of the teleosemantic theory, and found that it has the right status to answer the questions about intentionality which are the concern of this essay.

The only conclusion that could possibly follow from our progress up until now, is an endorsement of Millikan's teleosemantic theory. The theory has provided answers to

the questions about intentionality which opened this essay, as has been shown in this final section. However, this is not to say that we can conclude that Millikan has produced the correct theory of mental representation. We have not considered enough of the fine-detail of her theory to draw this conclusion, and neither have we answered objections to the theory in enough depth; we have often been content to merely provide indications as to what adequate answers would look like. The theory of content that philosophers ultimately settle with, if there ever is such a theory, might not be Millikan's. But we have found good reasons to expect that if there ever is such a theory, it will invoke teleological considerations based upon the evolutionary significance of the consumption of representations. For such considerations offer genuine insight into the nature of intentionality, and must surely enter into the final story to be told about intentionality, if there ever is one.

References

Adams, Fred & Aizawa, Ken (1994) "Fodorian Semantics" in Stephen Stich and Ted Warfield (eds.) *Mental Representation*, pp. 221-242. Oxford: Blackwell (1994).

Block, Ned (1978) "Troubles with Functionalism" in *Minnesota Studies in the Philosophy of Science*, 9, pp. 261-325.
Block, Ned (1986) "Advertisement for a Semantics of Psychology", in Stephen Stich and Ted Warfield (eds.) *Mental Representation*, pp. 81-141. Oxford: Blackwell (1994). Originally in *Midwest Studies in Philosophy,* 10: *Studies in the Philosophy of Mind*, pp. 615-678.
Block, Ned (1987) "Functional Role and Truth Conditions", in *Proceedings of the Aristotelian Society*, Supp. Vol. 61, pp. 157-181.

Braddon-Mitchell, David & Jackson, Frank (Unpublished) "The Teleological Theory of Content".

Brentano, Franz (1874) *Psychology from an Empirical Standpoint* (trans. A.C. Rancurello, D.B. Terrell, and L.L. McAlister). London: Routledge (1995).
Brentano, Franz (1930) *The True and the Evident* (trans. Roderick Chisholm, Ilse Politzer, and Kurt Fischer). London: Routledge (1966).

Chisholm, Roderick (1957) *Perceiving: A Philosophical Study.* Ithaca, NY: Cornell University Press.
Chisholm, Roderick (1958) "Intentionality and the Mental" in *Minnesota Studies in the Philosophy of Science*, 2, pp. 510-539.
Chisholm, Roderick (1960, ed.) *Realism and the Background of Phenomenology.* Atascadero, CA: Ridgeview.

Crane, Tim (1995) *The Mechanical Mind.* London: Penguin Books.
Crane, Tim (1998) "Intentionality as the mark of the mental" in A. O'Hear (ed.) *Contemporary Issues in Philosophy of Mind.* Cambridge: Cambridge University Press.

Cummins, Robert (1989) *Meaning and Mental Representation.* Cambridge, MA: MIT Press.

Davidson, Donald (1974) "Belief and the Basis of Meaning" in *Synthese*, 27, pp. 309-323.
Davidson, Donald (1987) "Knowing One's Own Mind" in *Proceedings and Addresses of the American Philosophical Association*, 60, pp. 441-458.

Davies, Martin (1995) "Intentionality" in A.C. Grayling (ed.) *Philosophy: A Guide Through the Subject*, pp. 275-301. Oxford: Oxford University Press.

Dretske, Fred (1981) *Knowledge and the Flow of Information.* Cambridge, MA: MIT Press.
Dretske, Fred (1983) "Précis of *Knowledge and the Flow of Information*" in *Behavioural and Brain Sciences*, 6, pp. 55-63.
Dretske, Fred (1986) "Misrepresentation" in Stephen Stich and Ted Warfield (eds.) *Mental Representation*, pp. 157-173. Oxford: Blackwell (1994). Originally in Radu Bogdan (ed.) *Belief: form, content and function*, pp. 17-36. Oxford: Oxford University Press.

Flanagan, Owen (1992) "Conscious Inessentialism and the Epiphenomenalist Suspicion" in Ned Block, Owen Flanagan, and Güven Güzeldere (eds.) *The Nature of Consciousness.* Cambridge, MA: MIT Press (1997). Originally in Flanagan's *Consciousness Reconsidered.* Cambridge, MA: MIT Press.

Fodor, Jerry (1975) *The Language of Thought.* New York: Crowell.
Fodor, Jerry (1987) *Psychosemantics.* Cambridge, MA: MIT Press.
Fodor, Jerry (1990) *A Theory of Content and Other Essays.* Cambridge, MA: MIT Press.
Fodor, Jerry (1994) *The Elm and the Expert.* Cambridge, MA: MIT Press.

Fodor, Jerry & LePore, Ernest (1991) "Why Meaning (Probably) Isn't Conceptual Role" in Stephen Stich and Ted Warfield (eds.) *Mental Representation*, pp. 142-156. Oxford: Blackwell (1994). Originally in *Mind and Language*, 6, pp. 328-343.

Frege, Gottlob (1892) "On Sense and Nominatum" (trans. Herbert Feigl) in Herbert Feigl, and Wilfrid Sellars (eds.) *Readings in Philosophical Analysis* (1949), pp. 85-102.

Grice, H.P. (1957) "Meaning" in *Philosophical Review*, 66, pp. 377-388.
Grice, H.P. (1989) *Studies in the Way of Words.* Cambridge, MA: MIT Press.

Harman, Gilbert (1982) "Conceptual role semantics" in *Notre Dame Journal of Formal Logic*, 23, pp. 242-256.

Heidegger, Martin (1927) *Being and Time* (trans. John Macquarrie and Edward Robinson). Oxford: Blackwell (1962).
Heidegger, Martin (1975) *The Basic Problems of Phenomenology* (trans. Albert Hofstadter). Indiana: Indiana University Press (1982).

Husserl, Edmund (1929) "Phenomenology" (trans. C.V. Solomon) in Roderick Chisholm (ed.) *Realism and the Background of Phenomenology.* Atascadero, CA: Ridgeview (1960). Originally in *Encyclopaedia Britannica*, 14th Edition.

Jackson, Frank (1997) Reply to: "Naturalism and the Fate of the M-Worlds" by Huw Price, in *Proceedings of the Aristotelian Society*, Supp. Vol. 71, pp. 269-282.

Lewis, David (1980) "Mad Pain and Martian Pain" in Ned Block (ed.) *Readings in the Philosophy of Psychology*, Volume 1, pp. 216-222. Cambridge, MA: Harvard University Press (1980).

MacDonald, Graham (1989) "Biology and Representation" in *Mind and Language*, 4, pp. 186-200.

Martin, C.B. & Deutscher, Max (1966) "Remembering" in *Philosophical Review*, 75, pp. 161-196.

Merleau-Ponty, Maurice (1945) *Phenomenology of Perception* (trans. Colin Smith). London: Routledge (1962).

Millikan, Ruth (1984) *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
Millikan, Ruth (1991) "Speaking up for Darwin" in G. Rey and B. Loewer (eds.) *Meaning in Mind: Fodor and his critics*, pp. 151-164. New York: Blackwell.
Millikan, Ruth (1993) *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Papineau, David (1987) *Reality and Representation*. Oxford: Blackwell.
Papineau, David (1993) *Philosophical Naturalism*. Oxford: Blackwell.
Papineau, David (1996) "Discussion of Christopher Peacocke's *A Study of Concepts*" in *Philosophy and Phenomenological Research*, 56, pp. 425-432.

Parfit, Derek (1984) *Reasons and Persons*. Oxford: Clarendon Press.

Peacocke, Christopher (1992) *A Study of Concepts*. Cambridge, MA: MIT Press.

Putnam, Hilary (1975) "The Meaning of 'Meaning'" in *Minnesota Studies in the Philosophy of Science*, 7, pp. 131-193.
Putnam, Hilary (1987) *The Many Faces of Realism*. Illinois: Open Court.

Quine, W.V.O. (1960) *Word and Object*. Cambridge, MA: MIT Press.

Russell, Bertrand (1921) *The Analysis of Mind*. London: George Allen and Unwin Ltd.

Sartre, Jean-Paul (1940) *The Psychology of the Imagination* (trans. Bernard Frechtman). London: Routledge (1995).

Searle, John (1983) *Intentionality*. Cambridge: Cambridge University Press.
Searle, John (1992) *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
Searle, John (1997) *The Mystery of Consciousness*. London: Granta Publications.

Sellars, Wilfrid (1958) "Intentionality and the Mental" in *Minnesota Studies in the Philosophy of Science*, 2, pp. 507-509, and pp. 521-539.
Sellars, Wilfrid (1963) *Science, Perception, and Reality*. New York: Humanities Press.
Sellars, Wilfrid (1981) "Mental Events" in *Philosophical Studies*, 39, pp. 325-345.

Tye, Michael (1995) *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

Financial support from the Humanities Research Board of The British Academy made possible the production of this thesis.