SIX TRICKY PIECES:

NATURALISTIC ACCOUNTS OF MENTAL CONTENT AND THE PROBLEM OF DETERMINACY

CONTENTS

ABSTRACT

The present report identifies six problems concerning the determinacy of mental content which confront naturalistic accounts of such content. In the light of these six problems, an evaluation of representative exemplars of four alternative types of naturalistic account (namely basic causal, informational, producer-based teleological and consumer-based teleological) is provided. It is argued that whilst each type of account makes important contributions none satisfactorily addresses all six difficulties. The results of this evaluation are used to propose an alternative naturalistic account of mental content which promises to meet this challenge in a more satisfactory way. The proposed strategy can be seen as a blend of a consumer-based teleosemantics with an information-based condition. Hence, the account is broadly teleological and appeals: (i) both to the proper functions of representation producers and to the proper functions of representation consumers; (ii) both to the relations between an internal state and its external conditions and to the relations between internal states.

In section II the first five of the six determinacy problems are presented against the backdrop of the shortcomings of simple causal accounts of mental content. The problems which are presented here are: The horizontal problem (**HP**); the vertical problem (**VP**); the problem of inscrutability of reference (**PIR**); the problem of necessarily co-referring/co-extensional representations (**PNC**); the problem of empty representations (**PER**). The disjunction problem (**DP**) is presented as a special case of (**VP**).

Section III turns the attention to more complex causal accounts with a focus on the informational semantic theories of Dretske (1981) and Fodor (1987, 1990b). In section 3.1., a detailed exposition of three distinct notions of information is provided so as to: (i) establish the foundation for the critical discussion of the theories which follows; (ii) establish the basis for the notion of information which I develop as a part of my blended theory in section VI. Section 3.2.1. then presents a generic information-based account and section 3.2.1.1. examines Dretske's development of this approach in attempting to deal with the first five

indeterminacy problems. It is suggested that though the solutions proffered are intuitively attractive, his account is ultimately unsuccessful in this task.

In terms of **HP** a line of argument is developed which will recur, in adapted forms, until the end of section 4.1.: Although information-based accounts give us more discriminative machinery in addressing the horizontal problem they face crucial difficulties in excluding disjunctions of proximal features from the mental contents ascribed. It will be suggested that the only way in which an advocate of a purely informational semantics could successfully deal with this problem is if he could show that such proximal disjunctions are open, and, hence, not projectable. However, it will be shown that the arguments for this conclusion are wanting. The line of argument which this section establishes, therefore, is crucial to motivating the adoption of a consumer-based teleosemantic background theory on which to graft an informational content condition.

In terms of **VP**, I will attempt to show that Dretske's (1981) account is a sophisticated relative of a Type 1 theory of content. I will proceed to show that it succumbs to the standard line of criticism of Type 1 theories (i.e. that they are either false of circular) despite this sophistication. This discussion will, therefore, establish a line of criticism which will recur in section 4.1. and will help to justify the abandonment of any Type 1 theory in favour of a consumer-based teleosemantics based on the theory presented in section 4.2.

In terms of **PNC**, **PIR**, and **PER**, I will consider and reject one of the two most plausible ways in which a pure informational semantics can attempt to deal with these problems, namely by appeal to the distinction between complex and simple concepts. In part this will be achieved by implicitly suggesting that how a representation is acquired and the relations which it bears to other representations are central elements of which to take note in developing an adequate solution to **PNC** and **PIR**. This theme will be continued in the discussion of Dretske's approach to **PER**. In particular, it will be suggested that the complex/simple distinction cannot account for empty would-be natural kind terms, the meaning of which may *depend* (in a suitable sense) *on*, whilst not being *identical to* the

content of other representational structures. Hence, the rejection of this line of response to **PNC**, **PIR** and **PER** will contribute to the motivation for accepting the response to these three problems which is defended in section 4.2.

The discussion of Fodor will attempt to show that his asymmetric dependence theory can neither adequately resolve the problem of proximal disjunctive contents nor adequately resolve **DP**. The central line of argument will be that in order for Fodor's proposal to work he must illegitimately prioritise non-disjunctive over disjunctive potential contents. Thus his account is either false or circular. This then completes the discussion of the approach of pure informational accounts to **HP** and **VP** and sets the stage for the adoption of a teleological element in part VI.

Fodor's (1994) approach to **PIR** is reviewed. It is suggested that its focus on the internal relations which a representation bears to other representations within a system is very valuable, even though the particular form in which Fodor presents it is ultimately untenable. The three deficits which the account displays are used to motivate a consumer-based teleological approach to the problem in section IV.

Section IV introduces and examines two teleological accounts of content (a producer-based and a consumer-based account). In section 4.1. it is argued that an appeal to teleology shows early promise in resolving the problems of proximal disjunctive contents and **DP**. However, with respect to **HP** this promise is undermined by a specific version of the sixth determinacy problem, namely the problem of complex causal roles (**PCR**). The rejection of Dretske's solution to this problem completes the demonstration, begun in section 3.2. that no appeal to the process of learning can salvage an informational semantics (so long as the focus remains with the representation producer, that is). With respect to **VP** the promise of a producer-based teleosemantics is undermined by the *general* problem of **PCR**.

It is furthermore argued that producer-based teleological accounts developed on the basis of an informational semantics face the fundamental problem that information is not necessary for representation within an evolutionary perspective. In section VI the blame is placed partly on the producer focus of such accounts and partly on the kind of conception of information which they adopt.

In section 4.2. a representative consumer-based teleological account is introduced, based on Millikan (1984). It is argued that such an account shows distinct promise of being able to develop the central value of an appeal to teleology in dealing with the problems of disjunctive proximal contents and **VP**/**DP**. However, with respect to **HP**, in order for this promise to be realised such an account must adequately deal with the problem that the contents generated on this account may not fall on the horizontal axis at all, resulting in unintuitive contents being attributed. With respect to **VP**, the realisation of the account's promise is also contingent on it being able to resolve the problem generated by the possible existence of multiple consumers.

The overwhelming strength of Millikan's account is in its application to **PNC**, **PIR** and **PER**. The details of this application are developed so as to argue, against the backdrop of the failings and the reasons for those failings, of the prior theories surveyed, that this sort of account can potentially, in principle, offer a satisfactory solution to all three of these problems. It is suggested that a crucial piece in the puzzle is Millikan's isomorphism requirement.

Section V introduces the sixth of the six determinacy problems, **PCR**, as a problem which potentially infects all teleological accounts. Four distinct types of response to **PCR** (those of Agar, Neander, Millikan and Papineau) are surveyed and it is suggested that though each makes valuable contributions to the adequate solution of the problem, none is ultimately satisfactory. This completes the explication of the virtues and vices of the particular accounts surveyed and sets the agenda for the account to be presented in section VI.

In this final section, the notion of information and information-based semantics is revisited, so as to develop an information-based condition which is compatible with a consumer-based teleosemantics. Ultimately it is argued that through an adapted application of Neander's proposed response to **PCR** to Millikan's consumer-based teleosemantics, combined with Papineau's focus on more complex systems we can generate a more satisfactory solution to the problem of complex causal roles. The addition of a broadly causal/informational condition to this version of teleosemantics is coherent, well motivated and promises to deal satisfactorily with the remaining problems of **HP**, **VP**, **DP**, the problem of multiple consumers and the problem of counterintuitive contents.

INTRODUCTION

In thinking about our own minds and those of others, in explaining behaviour and in reasoning about the world, we attribute a range of mental states which have a certain mental *content*. That is, our commonsense view of the mind credits a large class of mental states or events with the property of representing (or being able to represent) parts of the world in which that mind exists, and to represent those parts of the world as *being* a certain way. One central question for a theory of mental content, then, is: In virtue of what do mental states have this property of representing the world, of being *about* the world? This question arises in a particularly perspicuous form under three philosophical assumptions: (i) The Assumption of Content Realism (**CR**): That representational properties are objective features of a certain class of states (i.e. mental states) (see e.g. Lloyd, 1987); (ii) The Assumption of Content Externalism (**CE**): That the content of (at least some) representational states is determined, at least in part, by the relations (past or present) which such states bear to their external environment (see also Horgan & Kriegel, 2008); (iii) The Assumption of Psychological Naturalism (**PN**): That our best account of mental states and their representational properties should not posit or make appeal to entities or relations which are incompatible with (our most plausible version of) natural science (see also Field, 1978). In the light of the conjunction of these three assumptions our question, then, is: In virtue of what sort of external, naturalistic relation does an internal, physical state come to either have or to realise a representational property? This, hereafter, will be referred to as "The Central Question" (**CQ**).

Accounts of mental content which have attempted to address this question have characteristically found themselves confronted with a cluster of six problems concerning the *determinacy* of such content, i.e. the characteristic of mental representations to represent only a determinate subset of those objects to which they are related, and to represent those objects as having a determinate set of the properties (Lloyd, 1987). Successfully negotiating this cluster of problems consequently provides a crucial constraint on theory construction, and a crucial benchmark for theory evaluation. The present paper will critically examine representative exemplars of four alternative types of naturalistic account (namely basic causal, information theoretic, producer-based and consumer-based teleological accounts) in terms of their success in negotiating this cluster of determinacy problems. It will be argued

that whilst each type of account makes important contributions none satisfactorily addresses all six difficulties. The results of this evaluation are used to suggest a framework of principles within which an alternative naturalistic strategy could be developed in order to meet this challenge in a more satisfactory way. The paper ends with the construction and consideration of a naturalistic account of mental content constructed along these lines. The proposed account can be construed as a blend of a consumer-based teleosemantics and an information-based causal account. As such the account is broadly teleological and appeals: (i) both to the proper functions of representation producers and to the proper functions of representation consumers; (ii) both to the relations between an internal state and its external conditions and to the relations between internal states.

The project of providing a naturalistic account of mental content or mental representation is one of central importance in providing a materialistic theory of the mind (Field, 1978), compatible with a scientific outlook. A successful account of such content, therefore, would have importance for cognitive science and for any branch of philosophy which subscribes to a naturalistic outlook and requires a "theory of thought" (Godfrey-Smith, 2004). Progress in the attempt to resolve the difficulties revolving around the determinate nature of mental content which have troubled such accounts would, therefore, be a significant advance in the overall project. The present paper attempts to lay the foundations for the development of one way in which such progress may be achieved.

I

PRELIMINARIES

1.1. The Representation Relation and its Relata

Fodor (1984) suggests that a theory of representation (in general) should, at least, deliver us necessary and sufficient conditions, C, to complete the following formula:

(R) R represents S iff C.

In the context of **CQ** I take 'R' to range over mental states specified in a topic neutral (see Armstrong, 1980) fashion, i.e. to range over the set of internal states/states of affairs.

1.1.1. *Delimiting the Scope of 'S'*

The domain of 'S' includes states of affairs, represented by formulae of the form 'the x is F' (where x ranges over particulars and F over properties or structured sets of properties) (e.g. Dretske, 1981, 1986; Stampe, 1977). With this as the domain of S, articulating conditions C would provide an analysis which could serve as the basis for a naturalistic reduction of the content of simple, perceptual beliefs with 'x' serving as an indexical expression (e.g. the belief that *that* is a horse) (cf. Dretske, 1981). Under certain further assumptions, however, at times we may wish 'S' to range over particulars or properties. The most prominent of these assumptions are that a satisfactory theory of content will: (a) include a combinatorial semantics, and (b) account for more simple representational abilities (e.g. the representational abilities of animals which are cognitively more simple than humans).

Accounts of mental representation, therefore, can differ in terms of whether they adopt either of these assumptions, and, therefore, whether they take the fundamental unit of analysis (i.e. the domain of S) to be entire propositional attitudes (e.g. beliefs and desires), or whether they take this unit of analysis to be the set of semantic elements out of which such attitudes are composed.

In the present paper I will proceed, for the most part, on the basis of the assumption that there are mental representational states which express only properties and that there are such states which refer only to individuals (hereafter **ACA**). I will proceed in this way for three reasons. The first is that assumptions (a) and (b), above seem to me to be *prima facie* defensible, as does the claim that (a) and (b) together support the idea that the domain of S should include not only states of affairs but properties, individuals and relations as well. This, of course, requires that we accept a version of Realism about Universals, and this latter view is one which should, to my mind, admit of defence independent of considerations concerning mental or linguistic representation. Hence, the assumption adopted here is contingent upon the possibility of such a defence.

Since (a) and (b) and **ACA** will implicitly inform the conclusions presented in the present paper I should note the main reasons for my acceptance of them. The reason for my acceptance of (a) is the standard consideration that the existence of a combinatorial semantics of mental representations accounts best for the *productivity* and *systematicity* of certain sorts of representational states (like beliefs) (see, e.g., Fodor (1987) and Fodor and Pylshyn (1988)). The former refers to the property of the class of belief states which are in principle possible (under appropriate idealising assumptions which abstract away from certain sources of limitation on human powers of computation, e.g. limited resources) to be indefinitely large (Fodor & Pylshyn, 1988). The latter refers to the property of there being an intrinsic, but contingent, link between the ability to think certain thoughts and the ability to think certain others, at least for suitably demarcated sets of thoughts (Fodor, 1987; Fodor & Pylshyn, 1988) (e.g. the ability to believe that the dog bit the man and the ability to believe that the man bit the dog).

The reason for my acceptance of (b) is that it is intuitively plausible to attribute representational capacities to organisms which are simpler than humans, particularly given the assumption that mental representation is a *natural* phenomenon (cf. Godfrey-Smith, 2002). However, first it is doubtful whether such organisms have the ability to possess complex representations with subject-predicate structure, or whether such types of

representations are required in order to explain the behaviour of such organisms. Second, it is plausible to think that levels of natural complexity tend to occupy a continuum, 'shading off' at the edges (Godfrey-Smith, 2002). It therefore seems equally plausible to think that levels of cognitive complexity, including the complexity of mental representations, also admit of some variation in degree. This suggests, though doesn't show decisively, that we would (i) allow for the sort of complexity of mental content evident in complex cognitive systems such as human beings; (ii) allow for the lesser complexity of representations of simple states of affairs; (iii) allow for the simpler representation of properties and particulars on their own.

The second and third reasons for adopting **ACA** are methodological. First, this treatment accords well with the fact that the six determinacy problems to be discussed in the present paper are most clearly articulated as concerning those features of representational states which are predicate structures. Second, this treatment seems to be fair to the primary exemplars of naturalistic theories which will occupy the focus of the first five chapters of the present paper. This is more or less explicit in both Fodor (1987, 1990a, 1990b) and Dretske (1981, 1988) Millikan (1984) takes her theory to be an attempt at accounting for *all* mental representation. Nevertheless, as her discussion of various examples in the literature attests, in making the dialectic between her approach and that of, e.g., Fodor or Dretske perspicuous her account can certainly be *taken* as an account of those aspects of representational states which express properties (i.e. the predicate-like features of such states).[1]

It must be emphasized that one needn't adopt the assumption that mental representations admit of a combinatorial semantics in order to accept the idea that a complete naturalistic semantics should account for representational states which express only properties. One might, for example, think that whilst representational capacities such as beliefs are non-compositional, such capacities exist only in higher mammals, even though simpler representational abilities can be found elsewhere in the natural world. It should also be emphasized that whilst the aforementioned assumptions are adopted in the background for the purposes of the present paper, no criticism will be made of theories on their basis. In the case

[1] See, for example, her (1989) discussion of Dretske's (1986) magnetosome example.

of non-combinatorial accounts of propositional attitudes (e.g. that proposed by Papineau (1993) (see section 6)), criticism will be made of the non-combinatorial nature of the theory only insofar as this hinders either the successful achievement of the goals of such an account or the successful resolution of the problems of the determinacy of mental representation.

In what follows, unless matters turn on the distinction, I will simply use boldface capital letters to range over types (including properties and types of states of affairs) and capital regular-face letters to refer to tokens of those types (including particular individuals and particular states of affairs).

1.1.2. *Two Aspects of the Relation '…represents…'*

A complete theory of mental representation (and a complete answer to **CQ**) would be required to specify not only the conditions on a representational state having a certain *content*, but also the conditions on an item qualifying for representational *status* (see Neander, 2004). The present paper will focus on the question of the determinants of the *content* of a representational state, rather than on the determinants of the *status* of a state as representational. As such I will not be concerned with whether the theories to be considered below adequately differentiate the class of representational states from the class of non-representational states. Furthermore, I do not intend my own positive proposals in the final section of the paper to be taken as aimed at this task. On this approach, therefore, a successful naturalistic theory of mental content would *not* be a complete naturalization of the class of contentful representational states. It would, however, be an adequate naturalization of the *property of intentionality* possessed by such states.

In order for this approach to be justifiable it must be shown that we can legitimately separate the question of representational status from that of representational content, at least in the sense that we needn't evaluate an answer to the latter question in terms of whether it adequately differentiates the representational from the non-representational.

This approach has been questioned, sometimes on the basis that unless we provide an account which determines both the status of a physical state as a representation and the representational content of that state at once, we will be confronted with the unpalatable consequence of having a state which qualifies as a representation, without being a representation *of* anything (Millikan, 1989).

However, this seems to me to place us in a false dilemma. Accounting for both aspects of representation in terms of a single set of determinants would be *sufficient* for avoiding the aforementioned unpalatable consequence. However, it does not seem to be necessary. First, one might maintain that the adequate set of content determining conditions is a proper subset of the adequate set of representational status determining conditions (see also Neander, 2004). In this way, one secures the result that a physical state cannot qualify as a representation without qualifying as a representation of something in particular.

It is, of course, apparently analytic that if a physical state qualifies as a representation of *p* (for some value *p*) then that state must qualify as a representation. However, as long as we are willing to forego the aim of providing a complete naturalistic account of representation for the more modest (though very important) aim of providing a complete naturalistic account of one aspect of representation (i.e. the property of intentionality), this needn't worry us. For, the more modest project does not aim to specify conditions on a *physical state* having a specific representational content. Rather, it aims to provide conditions C* so as to satisfy the following formula:

(R*) Representational state, R, represents S iff C*.[2]

Second, one might hold that although the adequate set of content determining conditions are *logically* independent of the adequate set of representational status determining conditions,

[2] The more modest project's aspirations to comprising part of a complete naturalistic account of representation depend, therefore, on the promise of a naturalistic reduction of representational status.

given the nature of each set of conditions, there is a *metaphysical* or *nomological* necessity which precludes the possibility of a state qualifying as a representation though not of anything in particular. I will not explore this possibility in detail here. However, one can see at least the outlines of how such an argument might go.[3]

1.2. *Aims of a Naturalistic Theory of Content*

It was claimed in the previous section that a complete theory of content should provide necessary and sufficient conditions, C, to complete formula **(R)**. This aim was then delimited to that of providing necessary and sufficient conditions, C*, to complete formula **(R*)**. However, at least some of the theories discussed below (e.g. that of Fodor (1987, 1990b)), aim only to provide *sufficient* conditions for a representational state to have a certain content (rather than another). In the case of other theories (e.g. teleosemantic theories) it is far less clear how they are intended to, and how they should, be construed (see, e.g. Braddon-Mitchell & Jackson, 1997; cf. Papineau, 2001).

In addressing this issue, we must distinguish two aims which one might have in tackling the problem of developing a naturalistic theory of mental content: (i) to show how it is *possible* for a wholly physical system to have representational properties; (ii) to discover the naturalistic basis of representation in some *pre-identified representational system* (e.g. in human beings). It is the first of these projects which I take to be the purely philosophical one; the second, is to a significant degree an empirical one.

There may be *many* ways in which states of a natural system can come to represent their world. In order to achieve aim (i), we need only to provide one of these ways; in order to achieve aim (ii) we need to provide that way which is not only possible within the framework

[3] For instance, depending on the details of the respective causal relations, holding a causal/functional role account of representational status, and a causal/informational account of representational content together may entail that no internal state which qualifies for representational status can fail to qualify as having some representational content or other (though whether this will be the intuitively correct content or not is another matter).

of a naturalistic metaphysics but which is also *possible* and *actual* given the nature of a specific natural system.

Of course for those that take the task of developing a naturalistic account of mental representation to be that of providing a *theoretical reduction* of intentional properties to properties recognized within a naturalistic metaphysics, aims (i) and (ii) collapse into one.[4] Presumably offering theoretical reductions is a respectable philosophical enterprise. Hence, simply claiming that the purely philosophical project of the naturalization of mental content is limited to aim (i), does not decide between whether philosophical theories of such content should be framed as providing both necessary and sufficient conditions or as providing only the latter.

For the purposes of this section, I will confine myself to a few general remarks which will be important to keep in mind in what follows.

Both the option of claiming to provide only sufficient conditions and the option of claiming to provide necessary and sufficient conditions bring with them distinct potential advantages as well as distinct theoretical burdens. Taking the first option apparently has the advantage that the burden of proof placed on the naturalistic theorist is lighter: Apparently, she needn't concern herself with demonstrating that paradigmatic cases of systems which possess representational capacities (e.g. humans) satisfy the conditions which she proposes, or that they do so in as general a way as our intuitive attribution of such states to these systems seems to require. Depending on the specifics of the externalist/naturalistic account of content which is provided, this option may also help to ease certain intuitive difficulties which apparently confront such accounts, such as the Swampman (Davidson, 1987) problem.[5] This

[4] See Kripke (1980) on the modal force of theoretical reductions.

[5] Swampman is a creature synchronically, qualitatively, identical to Davidson; that is, qualitatively, at the time of his coming into existence (say time t), identical to Davidson in every physical respect. However, by a freak of nature Swampman has his origins in a random materialisation of elements and organic matter found in a swamp. Hence, Swampman has no evolutionary or selectional history and when he first materialises he has had no causal interaction with the world in which he emerges. The difficulty is that, given that Swampman at t is physically identical to Davidson at t, and assuming, as we have done, a physicalistic ontology, surely his brain

is so, since such a theorist has, it seems, left the possibility open that Swampman has contentful states in virtue of satisfying some distinct set of sufficient conditions for content.

However, these two advantages are mitigated by consideration of the following two points. First, it seems difficult to see how such a theorist's claim that the proposed conditions are sufficient for *meaning/content* can be assessed independently of paradigmatic examples of contentful states. Second, so long as one claims that one's conditions are merely sufficient for a state to have a specific content, allowing that there may be distinct sets of such sufficient conditions, one leaves open the possibility that the very same system can independently satisfy more than one set of sufficient conditions. Unless, therefore, such conditions can be shown always to agree in the content attributed, the proponent of this strategy must forego claims to having decisively resolved the problem of indeterminacy of mental content (Boghossian, 1991). This is particularly pressing, if such a proponent wishes to *depend* upon the possibility of alternate sufficient conditions in dealing with problems such as Swampman.

On the other hand, the theorist who claims to be providing both necessary and sufficient conditions for a representational state having a specific content must commit himself to the claim that either paradigmatic representational systems satisfy his proposed conditions or else that such systems are not really representational systems at all. Presumably, this second alternative would be a highly undesirable one to take up on. However the first alternative commits the theorist to a substantial *empirical* claim, and this does not, *prima facie*, sit well with the purely philosophical methodology demanded by the *theoretical* question of how naturalistic mental representation is possible. More pressingly taking this alternative requires that the naturalistically inclined theorist must either provide a set of conditions for mental representation which accommodates the intuitions apparently marshalled by cases such as Swampman, or else must attempt to diffuse their force. Either way there is a substantial theoretical burden which such a theorist must face.

---

states and, in virtue of this, his mental states at t are identical to Davidson's at t? Yet the natural relations which Swampman bears to his external world seem to be completely different to those which Davidson bears to his. If content is fixed by certain such relations, then it seems that Swampman cannot have the same representational states as Davidson.

For the purposes of the present paper, I will, for the purposes of analysis, criticism and discussion, for the most part construe the various accounts considered as attempting to provide sufficient conditions only. Furthermore, this is how I intend my positive proposals in the final section to be construed. However, I adopt this line *tentatively* and in the full knowledge that the responsibilities entailed by the abovementioned theoretical burdens are ones that I cannot discharge in the present paper. To this end there is one important possible source of content indeterminacy which this paper cannot adequately resolve.

II

Basic Causal Accounts, Misrepresentation and the Determinacy of Mental Content

The most straightforward answer to **CQ** is that an internal state/state type represents some aspect of the world in virtue of the present or historical occurrence of actual causal interaction between that state/state type and that aspect of the world. One source from which such accounts may gain support is the consideration of the conjunction of: (a) the frequent differentiation between mental state content in linguistic terms (e.g. through the attribution of a propositional content to a belief/desire state), and (b) the popularity of causal accounts of the referential meaning of certain classes of linguistic items such as proper names (e.g. Kripke, 1980). Another source from which support may be gained is in the consideration of the arguments marshalled by, e.g., Putnam (1981) in support of **CE** (and, indeed those marshalled by Putnam (1975) and Burge (1979, 1986) in support of semantic externalism).

Even in this skeletal form, basic causal accounts highlight the first five of the six determinacy problems with which I will be concerned. Following Godfrey-Smith (1989) and Lloyd (1987) we can conceptualise the first two of these problems as comprising the horizontal and vertical axes of two-dimensional Cartesian space.

The so-called 'horizontal problem' (Godfrey-Smith, 1989, p.536) (hereafter **HP**) arises from the fact that what is represented in a mental state only features as *one part* in the causal chain which eventually leads to the tokening of an internal state type. In the case of perceptual beliefs, for instance, the perceived object reflects light in a certain way, causing a projection onto the retina, causing various electromagnetic impulses to eventually trigger an internal state in the subject's central nervous system. This chain of causes extends beyond the proper object of the representation, since this object has a particular causal history which accounts for its presence in the perceiver's environment. In virtue of what, then, is the object of an internal state one of these (more proximal/more distal) causes and not another? *Actual* causal histories (of the production of states) do not seem to provide enough discriminative machinery (Sterelny, 1990) to answer this question satisfactorily.

The, so-called, 'vertical problem' (Godfrey-Smith, 1989, p.536) (hereafter **VP**) arises from the need to differentiate the set of objects to which a mental state refers from, on the one hand, larger sets of which this set is a subset, and, on the other hand, smaller sets forming a subset of this set. The problem has different forms, but usually concerns the question of which aspect of an object the mental state represents. For example, if the content of the state is to be *rabbit* (and, hence, the state is to refer to the set of rabbits), this content needs to be differentiated from the content *leporida*[6] or *mammal*, on the one hand, and *rabbits with which one has had contact* or *this rabbit* on the other. We certainly want to allow for the extension of a mental representation to generalize beyond the set of objects with which one has had causal contact, and to generalize to a *specific* larger set, but simple causal accounts, in themselves, do not seem to provide a principled basis for this.[7]

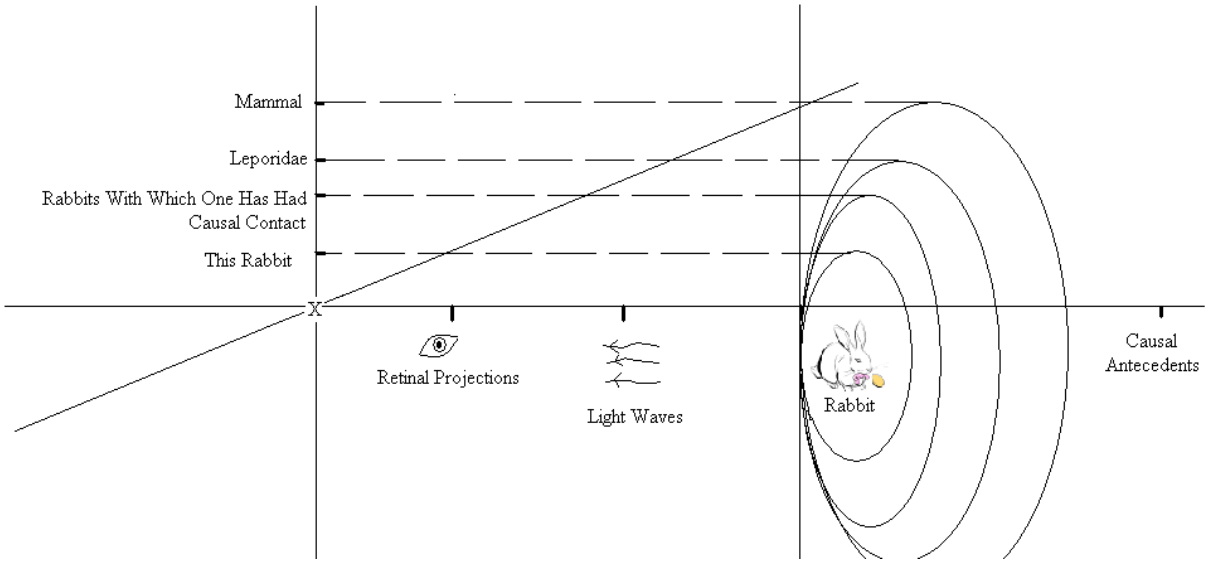

Figure 1

We can extend the Godfrey-Smith (1989)/Lloyd (1987) analogy by conceiving of the third type of indeterminacy problem as comprising a third Cartesian axis in three-dimensional

[6] Leporidae is the family to which rabbits belong.

[7] See figure 1 below, for a graphical representation of **HP** and **VP**.

space.[8] Basic causal accounts can be seen to be confronted by Quine's (1959) problem of the inscrutability of reference (hereafter **PIR**): The content *rabbit* refers to something of a certain ontological class, namely a unified object; the content does not refer to a temporal rabbit stage or to an undetached rabbit part. As causal contexts are extensional (see Davidson, 1967), however, and as these properties (hereafter 'Quine-type properties') are necessarily co-instantiated, every instance of a rabbit cause is an instance of a temporal-rabbit-stage cause is an instance of an undetached-rabbit-part cause. Any adequate account of mental content must account for the intensionality of contexts involving reference to certain types of representational mental states and, thus, for the fine-grained (Loewer, 2002) nature of content. Since causal contexts are *extensional* basic causal accounts on their own would seem be insufficient in this regard (cf. Shani, 2007).

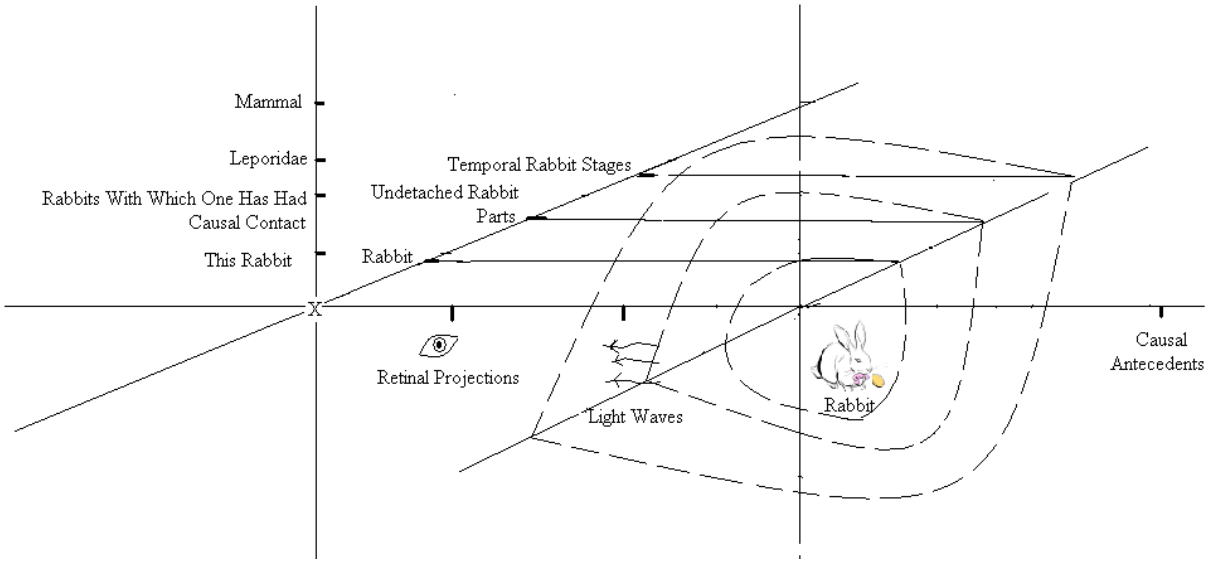

Figure 2

These three problems are really three manifestations of a general problem that all adequate accounts of mental representation must address, namely that of the possibility of misrepresentation (see, e.g., McGinn, 1982). That is, such accounts must show how it is possible for: (i) A state to represent an object, x, as F, when the object is not in fact F; (ii) a

[8] See figure 2 for a graphical representation of **PIR**. The lines demarcating the concentric ovals are perforated to emphasize that the classes of items demarcated are not proper subsets of one another.

state to represent the presence of a **Y** when no **Y** is present (Neander, 2004). An adequate solution to the three axes of the problem of determinacy, therefore, is intimately connected to an adequate solution to the problem of misrepresentation.

Adequately accounting for misrepresentation requires solving Fodor's (1984) problem referred to variously as the problem of disjunction, disjunctivism or (in lighter moments) disjunctivitis (hereafter **DP**). Basic causal accounts must accommodate cases of misrepresentation by making a distinction between types of internal states, **X**, and tokens of those types, X (Neander, 2004). Content is fixed for **X** due to causal interactions occurring between **X**s and **Y**s for some property or type of state of affairs, **Y**. Nevertheless, on occasion, a token of some **Z** (**Z≠Y)** may cause a tokening of **X**, whence misrepresentation. However, consider the union of the extensions of **Y** and **Z**; call elements in this union **YorZ**s. Since all **Y**s and all **Z**s are **YorZ**s, if there is an adequate mass of causal interactions between tokens of **Y** and tokens of **X** then there is an adequate mass of causal interactions between tokens of **YorZ** and tokens of **X**. Hence, in virtue of what do **X**s represent **Y**s rather than **YorZ**s? Of course if **X**s do represent **YorZ**s then since a **Z**, and hence a **YorZ**, caused a tokening of **X**, such a case is a case of *accurate* representation rather than misrepresentation, contradicting our original assumption. It will be noted that this is a special case of the vertical problem (Godfrey-Smith, 1989).[9]

[9] In this presentation of the disjunction problem I have used capitalized boldface letters to refer to elements in the union set of the extensions of **Y** and **Z**. However, the convention which has been adopted for the purposes of the present paper is to treat capitalized, boldface letters as ranging over types (i.e. properties and state of affairs types). Are we then to take '**YorZ**' as expressing a property?

In answer to this, it is important to note that there are at least two general ways in which the disjunction problem is presented in the literature. On the one hand, it is presented as a problem concerning the extensions of mental representational states (cf. Gates, 1996); on the other hand it is presented as a problem concerning the property expressed by a mental representation (cf. Fodor, 1990a, 1990b). Although it is uncontroversial that there are disjunctive predicates, it is substantially more controversial whether there are disjunctive properties. Assuming that there are no such properties, the disjunction problem is essentially the problem that it is indeterminate which of two (or three, etc.) properties a mental representation expresses, where at least one of these properties is picked out *in English* by a disjunctive predicate. On this reading, given that the property expressed by a predicate determines the extension of the predicate, the property expressed by '**YorZ**' will determine an extension which is disjunctive in the sense that it is the *union* of the extensions picked out by the two predicates used to form the disjunctive predicate 'Y or Z'. One apparent problem for viewing the disjunction problem in these terms is that it seems to license the view that the property which is expressed by '**YorZ**' is a property which is shared by **Y**s and **Z**s. However, it is not clear that for any two representationally relevant properties (that is properties which can feature in a naturalistic theory of content), there is always a representationally relevant property shared by every object which has either of the first two properties. The assumption which must

It has been alluded to above that mental representations are determinate not just in the sense of having a specific object (class of objects) but in representing the object in a specific way. The fourth and fifth determinacy problems concern this second aspect of determinacy. The fourth problem concerns the possibility of representational contents which are distinct yet necessarily co-refer (hereafter, **PNC**). For example, the belief that water is cool seems to be distinct from the belief that H2O is cool. Yet, since H2O *is* water (and necessarily so (Kripke, 1980)), every water causing is an H2O causing and so it seems that on a basic causal account, the content of these two beliefs must be the same.

The fifth problem is that of empty representations (**PER**). Basic causal accounts face difficulties in: (i) Accounting for states which apparently have content but do not have a referent, for on such theories world-brain actual causal interactions determine both referent and content; (ii) accounting for how such referentially empty representations could differ from one another in terms of content, since on such accounts all such representations are, presumably, contentless (see also Greenberg & Harman, 2005).

---

be made, therefore, is that since both **Y**s and **Z**s cause tokenings of **X**s, they must share some property in virtue of which they do this. But it is far from clear that this is true.

It may be better, therefore, to construe the disjunction problem as being concerned simply with the extensions of mental terms: That is the problem of disjunctivism is the problem that it is indeterminate as to what the extension of a mental representations is.

III

More Complex Causal Accounts, Indication and Information

Basic causal accounts attempt to fix content by appeal to *actual* causal relations and falter on the first five problems of determinacy. Increased discriminative machinery (Sterelny, 1990) would be supplied by appeal to relevant *counterfactual* cases as well. This is the central motivation behind complex causal accounts, including the informational/indicator semantics of Dretske (1981) and Fodor (1987) (Sterelny, 1984; Godfrey-Smith, 1989). The present section will discuss these accounts and their contributions to solving the first five indeterminacy problems. Section 3.1. presents three distinct versions of the notion of 'carrying information' which can be extracted from the work of Dretske (1981, 1986, 1988) and Fodor (1987, 1990a). Sections 3.2. then discusses the distinctive information-based accounts of mental representation which Dretske (1981) and Fodor (1987, 1990a, 1990b), respectively, develop and their contributions to the resolution of the problems of indeterminacy.

3.1. Indication, Information and Reliable Co-variation

3.1.1. The Accounts

Dretske (1981, 1984) attempts to develop a naturalistic theory of mental representation founded on a semantic account of information, informed by Shannon's (1948) mathematical theory of communication. On this account of mental representation the fundamental condition on a state's representing some aspect of the world is that that state carry information about that aspect of the world. I take the two central features of Dretske's account of information to be that:

(i) The information relation is determined by the relation of reliable indication.

(ii) The relation of reliable indication can be explicated in terms of conditional probability and natural law.

The indication relation, in turn, is best taken to obtain between types of states of affairs (Godfrey-Smith, 1992) (although Dretske (1986) restricts it to tokens) such that:

(I) **X** indicates **Y** iff if an **X** occurs then a **Y** occurs (Neander, 2004).

The successive developments of Dretske's account of information can be seen as different ways of interpreting the given conditional. Dretske's (1981) account of information requires that r's being **X**' (hereafter **X**(r), following Loewer (1987)) carries the information that **Y**(s) iff (i) the conditional probability of **Y**(s) on **X**(r) is 1, and (ii) this correlation is underwritten by natural law. In particular Dretske (1981, p.245, n.1) requires that there be a "nomic…regularity between these event types…which *nomically precludes* [**X**(r) when it is not the case that **Y**(s)]".

In order to avoid scepticism about the existence of information and to adequately handle his own paradigmatic examples of the information relation, Dretske (1981) suggests that the nomic regularities which underwrite the relation of carrying information should be understood as relativised to certain specifiable standing conditions. It is relative to these 'channel conditions' (i.e. that set of physical items, parts and processes which mediate the transmission of information (cf. Millikan, 2000)) that the relevant conditional probability is 1.

These channel conditions are meant to *reveal* rather than *eclipse* the dependency which obtains between the relevant event types. The occurrence of an **X** must, therefore, do some work over and above the standing conditions in precluding the possibility of s not being **Y**. Hence, letting the set of channel conditions be denoted by $\{C_i\}_{i\varepsilon[1,n]}$, we can formulate Dretske's (1981) account of information as follows (cf. Loewer, 1987):

(DI_1) **X**(r) carries the information that **Y**(s) if and only if there exist some strict law of the form '□((...&**CC**)→...)' (where '**CC**' is a complete description of the elements of $\{C_i\}_{i\varepsilon[1,n]}$, under the relevant types) which is true for the substitution of **X** in the first open position and **Y** in the second, and there exists no law of the form '□(**CC**→...)' which is true for the substitution of **Y** in the second position.

Dretske (1986, 1988) weakens this account by allowing that the relevant conditional probability may be less than one (though sufficiently close to it), and allowing that the correlation need not be underwritten by a causal or nomic relation. There must however be a contextually relevant condition which *explains* the existence/persistence of the correlation, thereby eliminating the possibility that it is a merely accidental/chance correlation (1988, p.57).

Given this interpretation of Dretske (1981), we must construe his (1986, 1988) versions of the notion of information as being based on the idea that there are cases in which one would intuitively allow that **X**(r) carries the information that **Y**(s), and yet for no set of conditions $\{C_i\}_{i\varepsilon[1,n],}$ is there a description **CC** such that $\Box((\mathbf{X}\&\mathbf{CC})\rightarrow\mathbf{Y})$ is true. In such cases, Dretske allows that as long as there is some condition which explains the correlation between **X** and **Y**, instances of the former can carry information about instances of the latter.[10]

The central notion tying all of these presentations together, therefore, is that the relation of 'carrying information that' is underwritten by a genuine, rather than accidental, correlation and dependence between the events related thereby. This dependence holds, moreover, only when certain counterfactuals of the form '**X**(r) would not be the case unless **Y**(s) were the case' are true (Dretske, 1986; McLaughlin, 1991). With this in mind, we can construe Dretske's (1986, 1988) presentation of information as depending on the idea that the presence of a condition which explains a correlation between two events can support counterfactual generalizations of the aforementioned form, even if this support is to a more limited degree than if such events were linked by a strict law relativised to channel conditions.[11]

[10] In what follows I will refer to this account of information as $\mathbf{DI_2}$.

[11] It should be noted that the conditions which obtain in the case of a non-accidental correlation between two events which are not linked by a strict law, and which are captured in an appropriate explanation of the correlation, differ significantly from the channel conditions. The latter are, whilst the former are not, such that it is possible for some description of them, '**C**', to be substituted for '**CC**' in '$\Box((\mathbf{X}\&\mathbf{CC})\rightarrow\mathbf{Y})$' to yield a true statement.

A somewhat different account of information is given by Fodor (1990a, 1990b). According to this account:

> **(FI)** **X**s carry information about **Y**s (under that description) if and only if there is a nomic connection between the property of being a **Y** and the property of being a cause of **X**s. That is, the generalization "**Y**s cause **X**s" is a law (1990a, p.57).

This shares with Dretske's (1981) analysis the feature that the relation of 'carrying information that' is underwritten by natural law in the very strict sense that there must be a natural law connecting **X**s and **Y**s (under those descriptions) if **X**s are to carry information about **Y**s (under those descriptions). However, it differs from Dretske's analysis in two (perhaps three) important ways. First, the direction of the dependence differs: **FI** allows, whilst $\mathbf{DI_1}$ does not that **X**(r) can carry the information that **Y**(s) even if **X**(r) can occur when it is not the case that **Y**(s), since an event type can have more than one cause.[12] Second on **FI**, though not necessarily on $\mathbf{DI_1}$ or $\mathbf{DI_2}$ the relation of 'carrying information that' is underwritten by a *causal* relation. Third whereas Dretske (1981) seems to require that the channel conditions be specifiable and that the relevant law once amended with these channel conditions be strict, Fodor allows that the laws in question may be *ceteris paribus* laws (Fodor, 1990a, 1990b; Boghossian, 1991).[13]

Hence, to summarise, there are a number of different positions one can take up in providing a naturalistic account of the information relation. These positions can be conceptualized on the basis of the following three questions:

(i) What is the direction of the dependence required?

(ii) Need this dependence be lawful?

(iii) Need this dependence be underwritten by the causal relation?

---

[12] Assuming that we can represent the lawful generalisation '**Y**s cause **X**s' as '$\Box(\mathbf{Y}\rightarrow\mathbf{X})$' (as Fodor (1987, 1990a, 1990b, 1991) seems to imply), we can see the difference between the two proposals immediately.

[13] Whether one takes this to be a genuine difference or not depends, however, both on Dretske's account of channel conditions as well as whether one thinks that the *ceteris paribus* condition can always in principle be filled out.

Different answers (and their permutations) to these three questions will determine a range of accounts of information. As we will see, utilizing these different accounts of information as the foundation for an account of mental representation will confront one with different problems.

3.2. Proposed Solutions to the Problems of Determinacy

The aim of the present section will be to examine the way in which information-based accounts of mental content attempt to deal with the first five problems of indeterminacy which were outlined in section II. The focus of the section will be the distinct proposals of Dretske (1981) and Fodor (1987, 1990b).

3.2.1. Informational Semantics

The generic idea lying behind information-based accounts of mental representation can be stated as follows:

> **(IS) X** represents **Y** only if **X**s carry information about **Y**s (at least, under certain circumstances).

Specific accounts in this family differ, however, in terms of which further conditions they stipulate as required for mental content.

3.2.1.1. Dretske (1981)

3.2.1.1.1. The Account

Dretske notes that on his account of information informational content will often not be unique. Given that **X**(r) carries the information that **Y**(s), **X**(r) will also carry all of the information *carried by* (or 'nested' in (Dretske, 1981)) **Y**(s). This is a significant *prima facie* drawback for using this theory of information as a basis for a theory of mental content which is able to cope with accommodating the determinate nature of such content. Dretske attempts

to address this problem first by utilizing the basis notion of informational content so as to generate an account of a semantic-like content which is determinate.

He proceeds via the following definitions. If **X**(r) carries the information that **Y**(s) only in virtue of **Y**(s) being nested in the further information *about s*, **Z**(s), then **X**(r) carries the information that **Y**(s) in analog form. If **X**(r) carries the information that **Y**(s) without carrying any further information *about s*, **Z**(s), in which **Y**(s) is nested then **X**(r) carries the information that **Y**(s) in digital form. If **X**(r) carries the information that **Y**(s) without carrying any further information *about any t (t not necessarily equal to s)*, **Z**(t), in which **Y**(s) is nested then **X**(r) carries the information that **Y**(s) in completely digitalised form. Hence, to carry **Y**(s) in digital form is to carry **Y**(s) as the most specific piece of information about s (Dretske, 1981, p.177). Any information which **X** carries in completely digitalised form is the most specific information carried by **X**. Since every item of information is nested in itself, this definition implies that if **X**(r) carries the information that **Y**(s) in *completely digitalised* form, then it carries it in digital form. However, the converse does not hold (Dretske, 1981, p.185).[14]

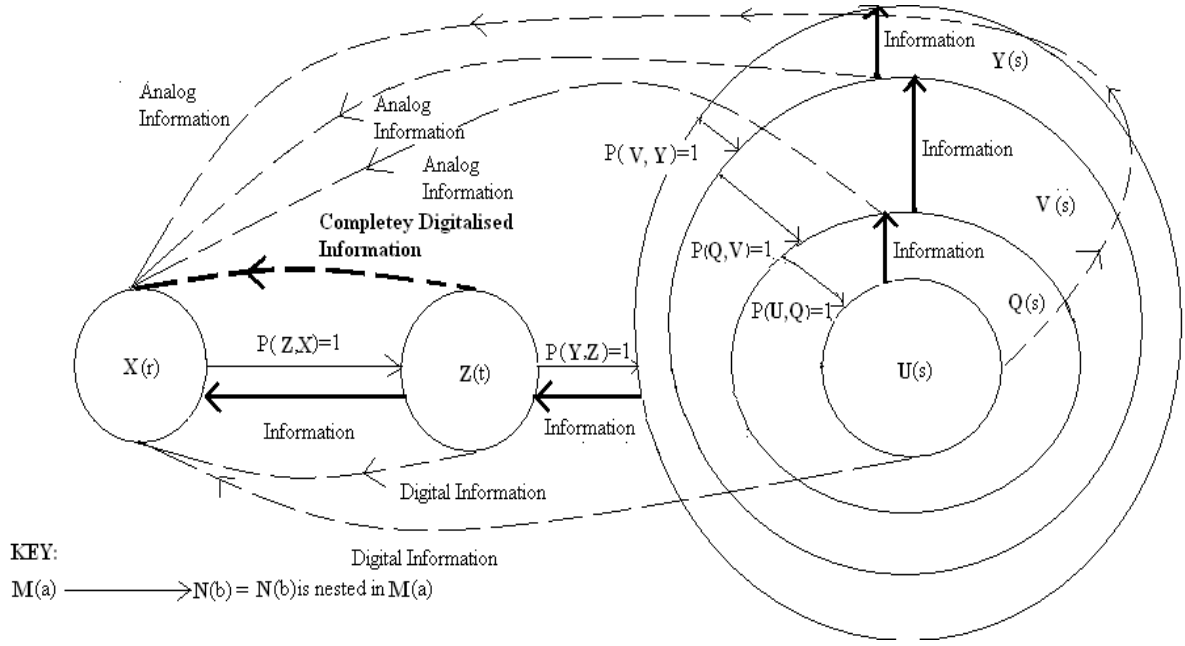

Figure 3

[14] See figure 3 for a graphical representation of the notions of information and analog, digital and completely digital forms of coding of that information. (cf. Dretske, 1981, p.177)

The notion of complete digitalization forms the basis for Dretske's notion of semantic content:

> **(SC)** For all r, s, for all **X**, **Y**, **X**(r) has **Y**(s) as its semantic content iff$_{def}$ **X**(r) carries the information that **Y**(s) in completely digitalized form. (cf. Dretske, 1981, p.185)

Now, **X**(r) has **Y**(s) as its semantic content only if **X**(r) carries the information that **Y**(s), and hence, only if s is **Y**. So identifying mental content with semantic content would leave no room for error or misrepresentation. In order to address this problem Dretske (1981, ch.8) distinguishes between two contexts in which **X** can be tokened: The learning and post-learning contexts. For any representational state, **X**, the content of the inner state type is that information which is completely digitalised by **X** within the learning context. Tokens occurring outside of this context, however, needn't carry this information, and when they do not misrepresentation occurs. Hence, the tokens of the type gain their representational content from the representational content of the type of which they are tokens, and the type gains its representational content from the information born by its tokens within a specific context.

Hence, the final account of mental content which Dretske (1981) gives is as follows:

> **(MC$_{D1}$)** For any **X**, **Y**, **X**(…) carries the information that **Y**(…)[15] iff$_{def}$ during the learning period, L, instantiations of **X**(…) came to have **Y**(…) as their semantic content.

[15] My understanding of Dretske's (1981, ch.9) account of concepts is that they are abstract meaningful structures which can be instantiated relative to objects indexically specified (i.e. the 's's, 't's, etc. of the preceding discussion) to yield *de re* beliefs. It is the concepts which provide the meaning specified by predicate clauses of the form '…is F'. The object of discussion here has changed from the representation of specific states of affairs indexically specified, to the abstract concepts which express that meaning which can be expressed in English through open sentences of the form '…is F'. Formulae of the form **Y**(…) are meant to convey this meaning.

The mechanism which Dretske envisions for the complete digitalization of the information that **Y**(…) involves a process of such information being used during L to shape the discriminatory and classificatory behaviour of the subject of the state tokens of **X**, assisted by training or feedback mechanisms. The role of learning and feedback in Dretske (1981) is simply that it is a *mechanism* for assisting in the development of a structure which carries a certain piece of information in completely digitalized form. It is the fact that this structure comes to have this piece of information as its semantic content which is determinative of content *not* any function rendered by the process of learning. This differentiates Dretske's (1981) account from his (1986, 1988) account.

3.2.1.1.2. The Proposed Solutions

HP

Such an account attempts to solve **HP** by: (i) Appealing to differences across nomologically possible worlds in the degree of correlation between more proximal and more distal features appearing in the causal production of a state; (ii) taking the proper object of representation to be that item, closest on the horizontal axis to the state, about which that state carries information (Dretske, 1981; Godfrey-Smith, 1989; Lloyd, 1987). (i) renders that set of properties which appear on the horizontal axis such that their occurrence is entailed (given suitable channel conditions) by the tokening of an **X**. (ii) renders that property, information about which has been completely digitalized by **X**.

This strategy is an intuitively attractive one. However, it is not clear that it is ultimately successful. The attempted solution rests on the idea that given the appropriate channel conditions, at certain points on the horizontal access there can be considerable variation in the properties which can be instantiated even though **X** is instantiated regardless. As such the conditional probability that any one of these properties is instantiated on the instantiation of **X** is less than 1. However, it seems to be plausible that the possibilities for variation in proximal intermediaries decreases as one moves toward the immediate causes of an internal state, so that at a suitably proximal distance from **X** on the horizontal axis, there will be no

room for variation of properties given the channel conditions. If this is so, then tokens of **X** will represent this proximal condition (since that is the piece of information which they carry in completely digitalized form). Hence, it becomes unclear how Dretske's (1981) account can yield distal contents (see Leon, 1987).

Arguably, the key assumption in the above argument is an empirical one. However, there is a more theoretical argument against this attempted solution of **HP** as well. Suppose that there is room for suitable variation at each point on the horizontal axis prior to but not including that point at which the property which is the intuitively correct object of representation, say **Y**, occurs. Suppose that for each point, *i*, between **X** and **Y**, the set of possible properties which can be instantiated without affecting the tokening of an **X** is $\{\mathbf{P_j}\}^{(i)}{}_{j\varepsilon[1,n(i)]}$. But then **X**s will carry the information that $\mathbf{P^{(i)}_1}(\ldots)$ occurred or $\mathbf{P^{(i)}_2}(\ldots)$ occurred or…or $\mathbf{P^{(i)}_n}(\ldots)$ occurred. In virtue of what is it the case that **X**s represent **Y**s rather than $\mathbf{P^{(i)}_1 V P^{(i)}_2 V \ldots V P^{(i)}_n}$? Given that the disjunctive predicate exhausts all of the possibilities, the conditional probability of the state of affairs type characterised by this disjunctive predicate on **X** is 1, and since it appears closer on the horizontal axis to **X** than does **Y**, it is a more viable contender for the information carried in completely digitalized form by **X**. That is, it is unclear why we cannot take the disjunction of possible proximate causes between the purported represented item and the internal state to be what is represented, rather than the purported represented item (Lloyd, 1987).[16]

[16] See figure 4 for a graphical representation of this point. In the figure the i-th position is conceived of as corresponding to the retinal projections of the rabbit, and the (in this model, finite) number of $\mathbf{P^{(i)}}$s are conceived as the various ways that the light reflected from a rabbit can strike the retina whilst still causing a tokening of an **X**.

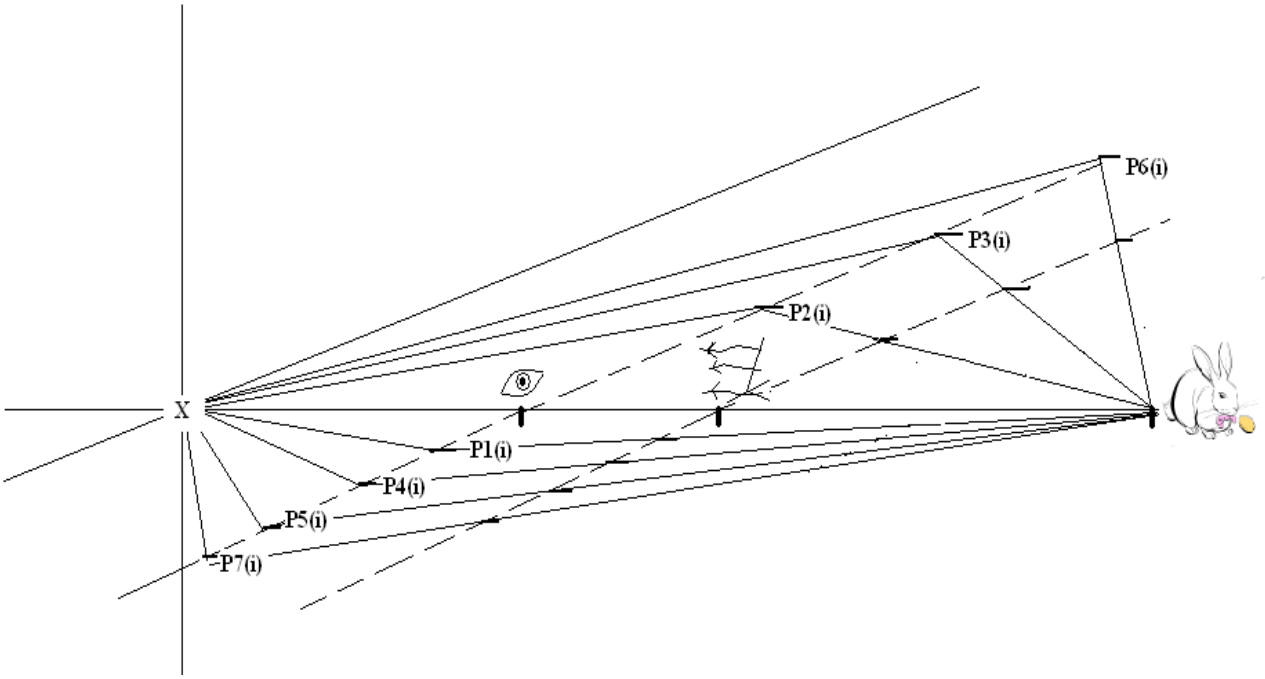

Figure 4

Two possible replies to this argument should be considered. First, it might be suggested that the requirement that there be a law which underwrites the relevant conditional probability might help one to avoid the difficulty. For, not every predicate is projectable, even if it is constituted entirely out of projectable predicates. However, this reply seems to be of no help to Dretske, since it depends upon there being available a theory of projectability which makes no appeal to any intentional or semantic notion, and some prime candidates for the title of 'correct theory of projectability' (e.g. Goodman's) would not meet that standard (Gates, 1996). But moreover, we have seen that the really fundamental notion in Dretske's theory of, the notion which the nomological condition is meant to capture, is that the relevant dependencies be counterfactual-supporting, and, hence, non-accidental. However, in these sorts of cases, given that the disjunctive predicates exhaust all of the relevant possibilities, it does seem that the relation between **X**s and the state of affairs type picked out by such a predicate is counterfactual-supporting. For, were an **X** to occur, such a state of affairs type would have to occur, since it is only *by* its occurrence that an **X** *can* occur.

The second reply is that the assumption that there are only finitely many possibilities for each *i* is unwarranted. After all, we are concerned with counterfactual possibilities here and there may be infinitely many for each *i*. It may be suggested that once we remove this assumption, the problem disappears because infinite disjunctions are not projectable (cf. Fodor, 1990b).

Well, surely the burden of proof here must fall on the shoulders of the proponent of **(MC_{D1})**, since it is he who is attempting to show that a satisfactory, wholly naturalistic account of mental representation is possible. Dretske (1986) and Fodor (1990b) have each offered arguments which attempt to discharge this burden. However, it seems to me that neither argument is successful.

Consider first the argument in Dretske (1986) presented in a somewhat different context. Dretske argues that in the case of more complex organisms, through the process of associative learning a number of different stimuli can come to trigger a single internal state. Dretske holds that in principle there is no upper bound on the number of different stimuli which can possibly come to fill this intermediary role. His explicit conclusion is that there is no *time-invariant* disjunction of state types which can be held to be the content of **X**. For present purposes, I will take Dretske's idea to be that although the class of proximal stimuli which *actually* will come to trigger tokens of the internal state type through associative learning is finite, the class of *possible* proximal stimuli which can come to trigger an internal state through associative learning varies over time and is infinitely large at *all* times.[17]

If this move is to do the work that Dretske requires of it then the every member of the set of possible proximal stimuli which can become triggers of **X** through associative learning must be such that it carries the information that **Y**(…). For otherwise, **X**(...) would not carry the information that **Y**(...). However, it is a substantial *empirical* assumption that this class is unbounded in size, one which is not supported by the fact that in the theory of classical conditioning the class of unconditioned stimuli can consist entirely of arbitrary members. For

[17] It may seem that this is not the most natural way to read Dretske (1986). However, it does seem to me to be the strongest way of reading his proposal which is in accord with the word of his (1986).

in standard cases of such associative learning the conditioned stimulus need not carry any information about the unconditioned stimulus.

More importantly, since content is fixed in the learning context (L), the relevant times, t, for which we need to consider the class of possible proximal stimuli which can become triggers of **X** through associative learning $\{ps_i\}_{i\varepsilon[1,\infty]}^{(t)}$, are just those within L. The idea being that though the channel conditions which obtain in L nomically preclude the occurrence of any non-**Y** at the same point on the horizontal axis as is occupied by **Y**, they do not preclude any of the members of $\{ps_i\}_{i\varepsilon[1,\infty]}^{(t)}$ from occurring. Hence, it would seem that mechanisms which underpin the processes involved in associative learning are all admissible within the learning context.

However, if the class $\{ps_i\}_{i\varepsilon[1,\infty]}^{(t)(j)}$ is specified as the class of all elements which can appear at point j on the horizontal axis, at time t, where this class is infinitely large due to the admissibility of the mechanisms of associative learning in L, then it seems that there would be nothing to preclude $\{ps_i\}_{i\varepsilon[1,\infty]}^{(t)(j)}$ from including *secondarily* conditioned stimuli.[18] Moreover, these secondarily conditioned stimuli needn't carry the information that **F**(…), since their status as triggers of **X** does not depend on their presentation in the presence of **Y**s. However, if $\{ps_i\}_{i\varepsilon[1,\infty]}^{(t)(j)}$ does include such stimuli, then **X**(…) does not carry the semantic content that **Y**(…) since it is possible within L for an **X** to be tokened in the absence of a **Y,** since an X will be tokened in response to a secondarily conditioned stimulus.

Hence, this first line of argument for the conclusion that there are only finitely many possible alternative causes at each *i* on the horizontal axis is unsuccessful.

The second line of argument for this conclusion (given in Fodor (1990b)) suggests that, at least in the case of medium sized perceptual objects such as cows, the class of proximal

[18] A secondarily conditioned stimulus is a conditioned stimulus which comes to elicit the target response due to its association with another conditioned stimulus.

stimuli is infinite, since belief fixation has a holistic character (Boghossian, 1991): What proximal stimuli you believe to be associated with cows (smells, sounds, images, etc.) will help to determine what proximal stimuli can cause a token of a mental symbol representing cows. Since, *any* proximal stimulus can become a trigger of the relevant internal state token given a suitable set of background beliefs, the class of possible proximal stimuli is infinitely large.

However, I think that this argument is a non-starter in the present context. For, though we might claim that the apparent circularity of the appeal to background beliefs can be avoided by focussing on the underlying physical dispositions required to implement that set of beliefs, the same sort of physical complexity can be attributed at any point on the horizontal axis. That is, if it is true that theory can mediate which *proximal* stimuli come to trigger an internal representational state, then it is equally true that theory can mediate which *distal* stimuli can come to trigger the same internal state type. To suppose otherwise would be the beg the question by illicitly assuming that one distal state of affairs is the proper representational object rather than another. However, why then can we not eliminate the distal state of affairs which is intuitively represented by the internal state from the status of representational object?[19] Hence, Fodor's appeal is either circular or unhelpful.

Hence, it seems that Dretske's (1981) account does not adequately resolve **HP**.

VP

The proposal for addressing the problem of **VP** appeals to similar resources as those utilized in **HP**. Appeal is made to differences across nomologically possible worlds in the degree of correlation between an internal state type, **X**, and the properties which define the possible supersets and subsets into which the possible represented, **Y**, falls. Hence, a state type, **X**, has the content *rabbit*, rather that *leporida, mammal, rabbits with which one has had causal*

[19] See also Boghossian (1991).

contact, or *this rabbit* because the nomic dependence obtains between **X**s and rabbits rather than mammals, etc. (cf. Fodor, 1990b).

However, the problem of disjunction is more tricky and highlights a more general difficulty concerning the notion of information/indication. If that **X** represents **Y** requires that **X**(…) has **Y**(…) as its semantic content (and hence carries the information that **Y**(...) according to $\mathbf{DI_1}$) then error would seem to be excluded. Suppose then that **X** represents **Y** can allow for P(**Y**,**X**)<1 (i.e. suppose we weaken $\mathbf{DI_1}$ to $\mathbf{DI_2}$). Then for any purported **Z** (≠**Y**), occurrences of which can cause occurrences of **X**, **X** correlates strictly better with **YorZ** than with **Y**.[20] **X** therefore has a disjunctive content, and extending this argument renders error impossible. On either conception of information, therefore, error seems to be impossible.

It has already been noted above that Dretske (1981) attempts to deal with this problem and allow for error by distinguishing between the learning and post-learning environments. Hence, Dretske's account seems to be a Type 1 theory (Fodor, 1990a), since it apparently attempts to fix content by specifying a context in which only one sort of information can be carried by **X**. The standard line of criticism against Type 1 theories is that they either do not solve the problems which they are meant to or that they do so in a circular way. In the discussion that follows I will show that: (i) Dretske's (1981) account is a very sophisticated Type 1 theory; (ii) it succumbs to the standard line of criticism regardless.

Fodor (1984) has objected to Dretske's account that even if the learning context/post-learning context can be adequately drawn it will not succeed in solving the problem of **DP** (Fodor, 1984). **($\mathbf{MC_{D1}}$)** attempts to allow for error by fixing the content of a representation, **X**, according to the information, **Y**, which it carries within (L), and then allowing that it needn't carry the information that **Y** outside of (L). Hence outside of (L) **X** can be tokened in response to both **Y**s and **Z**s (where **Z**≠**Y**). But if this is true, then if a **Z** *were* to occur during the learning period then it *would* have caused a tokening of **X**. However, it is the

[20] That is, P(**Y**,**X**)<P(**YorZ**,**X**)≤1.

counterfactual dependence between **X**s and states in the world which determines the informational content which **X** carries. Hence, it is not true that **X** carries the information that **Y** in (L), since it is not true that an **X** would not occur unless a **Y** were to occur: An **X** would occur without a **Y**, namely in the presence of a **Z**. Rather if **X** carries any information in (L) then it must carry the information that **YorZ**. Hence, **X** has the content **YorZ** outside of (L) and, hence, tokens of **X** caused by **Z**s are accurate representations not misrepresentations.

A second objection to **(MC_{D1})** questions whether the distinction between the learning and post-learning environments can be drawn in a naturalistic way (Loewer, 1987; Fodor, 1987), and can be drawn sharply enough to support the distinction between truth and falsity (or accurate and inaccurate representation) (Fodor, 1984, 1987, 1990a).

There seem to be at least two concerns here. First (see, e.g., Fodor (1987)), there is the concern that there is no *principled* distinction between the period of time during which a representation is undergoing the shaping of its content, prior to its application being evaluable as true and false (or correct and incorrect), and the period of time during which the content has been shaped and the representation is only being applied.[21] The distinction between truth and falsity, correct or incorrect application, is an all or nothing affair; however, the distinction between the learning and post-learning environment is a matter of degree. Hence the latter distinction cannot be the basis for the former. Second (see, e.g. Loewer (1987)), there is the concern that it seems plausible to think that different concepts will have different learning periods and, hence, what counts as the learning period for one concept can only be specified with reference to the content of the concept. Hence, the learning/post-learning distinction threatens to be question-begging as far as the project of developing a naturalistic account of mental content is concerned.

[21] This worry is highlighted by the fact that it seems intuitive that we should allow it to be possible for a mental representation to change its content over time. This, moreover, presumably occurs during the time when a representation, the content of which has already been shaped, is now being applied in a way that can be evaluated as true or false (or correct and incorrect).

However, it seems to me that there is a way for Dretske to respond to these objections. Dretske is not explicit about what constitutes the learning period; however, his comments are suggestive. Based on his primary and central presentation of his proposed solution to the problem of misrepresentation (1981, p.193) I suggest that we can frame the following necessary and sufficient conditions for a period of time to count as a learning period which are in accord with the spirit of his account:

> **(LP)** A period of time $[t_1, t_n]$ constitutes a learning period for a mental representation, **X**(r) iff there exists some state of affairs **F**(s) such that:

(a) At t_1 **X**(r) codes in analog form the information that **F**(s), relative to channel conditions **CC**;

(b) At t_n **X**(r) codes, for the first time, in completely digitalised form the information that **F**(s), relative to channel conditions **CC**.

Notice that neither the analysandum nor the analysans of **(LP)** individuates the mental representation which is a candidate to have a learning period with reference to its content. Hence, it is now open to us to identify (as Dretske does) the content of the representational state as that piece of information the development in coding of which defines $[t_1, t_n]$ as a learning period for **X**.

Three points are notable about **(LP)** the first is that the demarcation between the learning period and the post-learning period is sharp: As soon as **X** digitalises *any* piece of information which it has carried in analog form in the past, the learning period ends and the content of **X** is that information which it has digitalized, whatever that may be. Second, it is also possible that **X** carries information other than **F**(s); if any of this information is carried at a later point in completely digitalized form then that piece of information becomes the content of **X**. Hence, in order that **X** has a determinate content, we must stipulate that the content of **X** is that informational content that has been completely digitalized in the most recent learning period. Third, it is not a requirement of **(LP)** that the channel conditions relative to which **X** carries the completely digitalized information which it does at the end of

(LP) would have to obtain *throughout* the period $[t_1,t_n]$.[22] Rather, the channel conditions which obtain at t_1 and t_n and relative to which **X**(r) carries **F**(s) in analog form at t_1 and in digital form at t_n can change and, hence, fail to obtain at other times t_j, jε(1,n). [23]

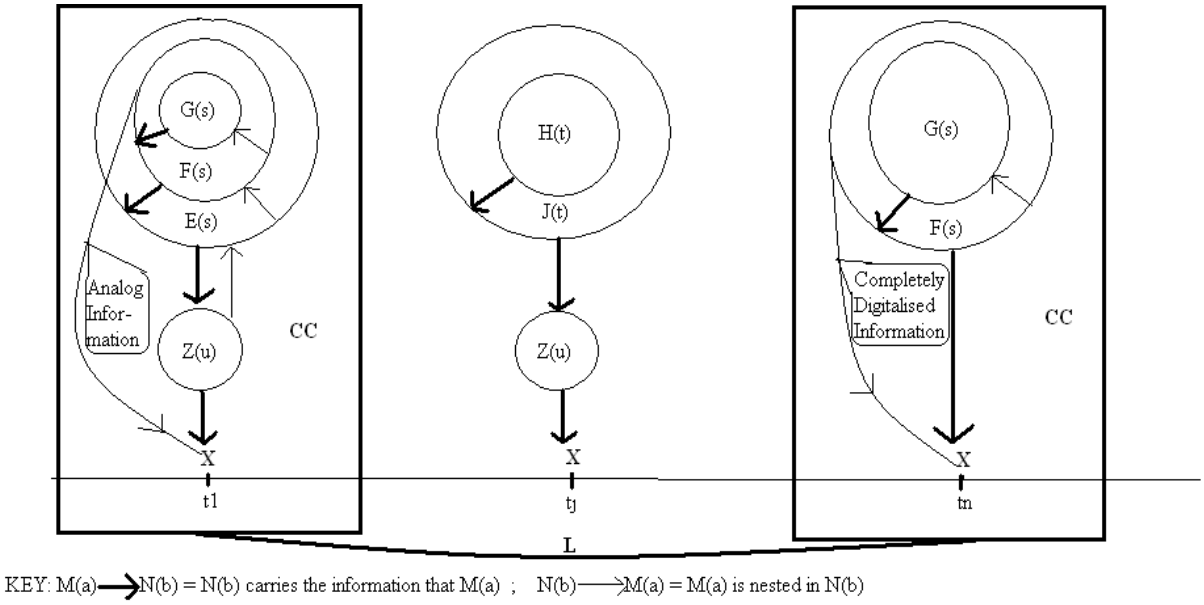

Figure 5

[22] **(LP)** has been constructed to secure this result, since there must be some mechanism which mediates this development of the ability to digitalise information. As Dretske himself implies, the most plausible mechanisms to fill this role involve the presentation of items from a *contrast class* which could easily elicit a token of the representational state concerned.

[23] See figure 5, below for a graphical representation of this construal of the learning period, (L). In the figure, F(s) might be an object's being a bird. This information is carried in analog form, since the object is a specific type of bird (e.g. a canary); this is represented by 'E(s)'. This piece of information is also carried in analog form since its relationship with **X** is mediated by some more proximal stimulus such as light striking the retina in a certain way. F(s) also carries further information, say that s is a mammal (this is denoted by G(s)). Over the course of time from t_1 to t_n the subject of S becomes able to 'generalize' over a large variety of particular ways in which birds reflect light, and to 'generalise, over a large variety of different specific kinds of bird. This may happen by the presentation at t_j of various non-canary birds (e.g. yellow budgies) and perhaps certain non-birds which are easily confused with birds. This latter possibility is represented by H(t) (notice that this information is also carried in analog form, since there must be a particular kind of non-bird presented, which may cause the same pattern of light to be projected onto the retina of the subject. Finally notice that **X** does not express **Z**, since there is no progression from analog to completely digitalized form.

Hence, **(LP)** is a naturalistic definition of the learning period which can support Dretske's project and which allows for: (i) a sharp enough distinction between learning and post-learning contexts; (ii) the change in meaning of a concept. It also demonstrates how we can allow that different concepts can have different learning periods, without our needing to specify the learning period with reference to the content of the concept.[24] Finally, it avoids Fodor's objection since so long as there is no pair of times (t_a,t_b) such that a,b ε [1,n], a≠1, b≠n, and **X** carries the information that **YorZ** in analog form at t_a and **X** carries the information that **YorZ** in completely digitalised form at t_b, then **X** has the content **Y** rather than **YorZ**.

It seems to me that IF the notion of channel conditions can be adequately articulated, there is nothing incoherent with attempting to salvage $\mathbf{MC_{D1}}$ by appeal to **LP**. However, the 'if' in the opening line of this paragraph is a big 'if' (literally and metaphorically), since it requires an account of channel conditions which are not only naturalistically specifiable and non question-begging but also non *ad hoc* (Neander, 2011). I don't have a knock-down argument that such a specification is not available; however, it is not at all clear that it is. Certainly *if* it is Dretske hasn't told us enough about channel conditions to see that it is. More likely, it seems that the problem of giving a naturalistic account of mental representation just recurs at the level of giving a naturalistic account of channel conditions. *This* is the fundamental problem which I see in the current proposed solution to **VP**, the problem of error and the problem of **DP**.

The threat of circularity is a problem which $\mathbf{MC_{D1}}$ shares, I would suggest, with Type 1 theories in general. In constructing a theory of content we want to isolate, as the *proper* relatum of the state, one of the many items to which a mental representation is related. However, since the representation *is* related to all of these items, it seems that *if* we can specify conditions under which the representation is related to one item only, then there are specifiable conditions for every other item to which it is actually related. Which of these

[24] Rather the specification proceeds in the opposite direction.

conditions are the content determining ones will depend, therefore, on which item is the desired or actual content of the state, rather than *vice versa*.

PNC and PIR

PIR and **PNC** receive a single treatment, by appealing to the distinction between complex and simple concepts. In terms of **PNC**, since the properties of 'being water' and 'being H2O' are identical, any state which has *water* as its semantics content has H_2O as its semantic content (Dretske, 1981, p.217). However, water beliefs can differ from H_2O beliefs by instantiating distinct concepts (Dretske, 1981; Gates, 1996) where in this case the concepts are differentiated in terms of degree of complexity: Whereas 'H2O' is a complex concept (comprised of the concepts 'Hydrogen', 'two' and 'Oxygen') 'water' is a simple concept (Dretske, 1981, p.217; Gates, 1996):[25] Whereas one can token **WATER** without tokening **Hydrogen**, one cannot token $\mathbf{H_2O}$ without tokening **Hydrogen**, since $\mathbf{H_2O}$ is (or instantiates) a complex concept (Dretske, 1981, p.264, n.2).

However, it seems to me that this attempted solution is ultimately unsuccessful.[26] One significant consequence of Dretske's attempted solution to **PNC** is that simple representational states which are equivalent in terms of semantic content must be identical (Dretske, 1981, p.217). However, we can generate cases in which two syntactically and *semantically*[27] simple representations have a felt difference in content.

[25] It is important to note that there really are two threads which come together in Dretske (1981). The one is that concepts which are equivalent in semantic content can be different in terms of their complexity. The other is that this difference in complexity results in a difference in *function* of the semantic structure. There are times at which Dretske seems to suggest that the difference in *function* is the constitutive difference between the relevant structures. In this section I'll deal with the first of these threads. The role of internal function in attempting to deal with these problems will be revisited in subsequent sections.

[26] The argument given below shows that the attempted solution does not work for predicate structures. Notice also, however, that even if it did it would have the theoretical disadvantage of precluding a unified response to the analogous problem to **PNC** which arises for name-like structures (a point noted by Fodor (1994), since these structures are in general syntactically simple.

[27] A semantically simple representation is one which is not equivalent to some description by which it is introduced. Semantically complex structures operate somewhat as placeholders for their equivalent descriptions. Hence, Dretske's type of solution can accommodate such cases since the underlying descriptions are composed of semantically different simples combined in ways which yield semantically identical complexes.

Consider Sam. Sam lives on the planet Zoot, which is very much like Earth in terms of certain of its atmospheric conditions. The inhabitants of Zoot, the Zootians, look very much like humans. The only difference is that the people in Sam's world fuel their bodies with a red viscous liquid, Raso, which takes the place in their lives of the items which we eat and drink on Earth. There is no water on Zoot and hydrogen and oxygen are very rare, so rare, in fact, that the existence of these elements is not known.[28] The Zootians' bodies can, however, absorb and process water just like our bodies (as they can digest the items which we eat on Earth, though they do not eat these items on Zoot).

Sam is a scientist on Zoot. One day, through a freak occurrence, a very small amount of water molecules forms on the electron microscope which Sam and his scientist friends use in their lab.[29] The sample is too small a quantity for Sam (or anybody else) to be able to see that the substance is clear or a liquid, or that it can satisfy any of the role descriptions which it does on Earth. Nevertheless Sam and a group of his scientist friends are able to experiment on the water molecules, noting characteristic effects and using ostension or the description of such effects to fix the reference of a term 'retaw'.[30] Over time a completely different group of Sam's scientist friends are able to separate the hydrogen and oxygen atoms in the water samples and investigate their properties. Again, through description or ostension they fix the reference of two terms 'Hydrogen' and 'Oxygen' respectively. Sam and all his friends get together for coffee to discuss their results and realize that retaw is H_2O.

Having worked so hard, Sam decides to take a well deserved vacation. The Zootians are scientifically a very advanced culture and have developed the requisite vehicles for high-speed space travel. Despite this no Zootian has ever visited Earth. Sam decides to hop in his spaceship and go where the Milky Way takes him. It so happens that Sam arrives on Earth.

[28] To make this plausible, we might suppose that there are atoms very similar to these which can perform a very similar role to that performed by hydrogen and oxygen on Earth.

[29] I'm not really sure whether one can 'see' a water molecule by using an electron microscope. The scientific details should be changed here if one can't.

[30] The term is taken from McGinn (1982), though is being use here somewhat differently.

Sam likes our planet and decides to stay. Since Sam is on vacation he avoids all contact with any scientifically literate Earthians. Over time, due to his interaction with members of communities of individuals who lack the concepts of Hydrogen and Oxygen Sam develops an intrinsically distinct internal structure which expresses the concept of water.[31] Sam, of course, never realizes that water is H_20 or that it is retaw. After some time Sam begins to get bored and so decides to explore the state of Earthian science. During this process Sam realizes that there is plenty of Hydrogen and Oxygen on Earth. One day whilst conversing with one of his newly made Earthian scientist friends he comes to realize that water is H_20, and hence that water is retaw. Sam regards these as substantial discoveries, very different to the discovery of the identity 'A is A'.

All of this has been framed in terms expressions in a natural language. However, the above seems to me to carry over *mutatis mutandis* to mental representation, given, that is, a suitable account of the mental representation of natural kind terms. However, such an account is presumably presupposed by the framing of the problem of **PNC**, and seems to be plausibly in the offing.[32] Hence, we have a case which I would suggest is intuitively correctly described as a case in which Sam's **RETAW** beliefs and Sam's **WATER** beliefs are different from one another. Yet, by the lights of informational semantics, each representation has the same semantic content. Moreover, there seems to be no difference in conceptual complexity between **RETAW** and **WATER**: Both seem to be conceptually simple. Hence, the complex/simple distinction on its own cannot provide informational semantics with a satisfactory response to **(PNC)**.

Dretske's response to **PIR** seems to be as follows. Rabbit beliefs (or mental representations), undetached rabbit part beliefs and temporal rabbit stage beliefs all have the same semantic content. However, these structures can differ from one another in terms of the complexity of the concepts involved, such that, e.g., it is not possible to have the second and third sorts of belief without also having concepts with the content *part* or *stage* (i.e. the concept involved is

[31] Hence, this acquisition is based on ostension or characteristic functional roles occupied by water on Earth.

[32] See the discussion of Fodor's asymmetric dependence theory below.

complex) whereas this is possible in the first case (i.e. the concept involved is simple) (1981, p.264, n.2; Gates 1996).[33]

As such, **PIR** is lumped together with **PNC**. Notice first, however, that the problem of **PIR** unlike that of **PNC** is a problem involving properties which, though they are necessarily co-instantiated, are non-identical and have *distinct* extensions (Gates, 1996). However, Dretske renders the semantic content of (and property expressed by) all representational states involved in **PIR** as the *same* (i.e. that property picked out by the disjunction (perhaps better, conjunction (see Dretske, 1984)) of every possible Quine-type predicate). Since the semantic content of these states is the same, by parity of reasoning with examples such as the water-H_20 case, it seems that we would be correct in believing that rabbits *are* undetached rabbit parts, that undetached rabbit parts *are* temporal rabbit stages and that temporal rabbit stages *are* rabbits (Gates, 1996, p.337). However, all of these thoughts seem to be patently false, though Dretske's account of semantic content and his attempted solution to **PIR** must, it seems, render them true.

There is an even more punishing problem for Dretske's proposal. A concept or cognitive structure will be complex if and only if the structure is composed in a certain way of discrete cognitive substructures (Dretske, 1981), such that the content of the complex is generated by the content of its parts and the manner in which those parts are combined (cf. Dretske, 1981, p.230). Simple cognitive structures are, according to Dretske, individuated only in terms of their semantic content. What, then, are the simple parts which combine (in the case of the complex representations appealed to in the solution to **PIR**) to express the property *rabbit&undetached-rabbit-part&temporal-rabbit-stage*?

[33] I must emphasise here that Dretske's treatment of Quine's examples is very brief, occupying only a single footnote of length one paragraph (1981, p.264, n.2). His treatment there purports to explain why "believing that *s* is a rabbit and believing that *t* (some part of *s*) is an undetached part of rabbit or the belief that *u* (some temporal cross section of *s*) is a temporal rabbit stage". Hence, Dretske takes it for granted that there is a naturalistic account to be had which can respect the distinction between referring to an object, some part of that object and some temporal cross-section of that object (though *de re* not *de dicto*).

Dretske gives two of them, namely a cognitive structure expressing the property *part* and a cognitive structure expressing the property *stage*. Hence, presumably the other constituents are cognitive structures expressing *undetached* (or perhaps just *detached*), *temporal* and ... *rabbit*! Hence, this proposed solution must *assume* that there is a cognitive structure with the determinate content *rabbit*. However, the problem of **PIR** is precisely in virtue of what there could be such a structure. Dretske's proposed solution, therefore, must, if it is to work, assume that the problem of **PIR** has already been solved. Hence, his proposed solution is really no solution at all.

PER

MC_{D1} entails that **X** has the representational content **Y** only if at the end of L **X** had the semantic content **Y**. However, if this is true, then at the end of L **X** carried the information that **Y**, and, hence, since the information relation is factive, at the end of L **Y** must exist.[34] However, empty representations are representations of non-existents. Hence, it seems that **MC_{D1}** cannot deal with **PER**.

Dretske seems to accept the above argument but limits its force to primitive representational states (i.e. those that are syntactically and semantically simple). Primitive representations cannot have empty contents (1981, p.229); however, complex representations can. Empty representations, such as representations which have the content which 'unicorn' does in English are, therefore, necessarily complex, built up out of simpler representational states which have non-empty extensions.

This seems to me to be the only response which Dretske can make to **PER** given his account of the information relation and the role of information in mental representation. It also seems to me to be highly intuitive to think that representations with empty extensions are complex

[34] Even on weaker notions of information than **DI_2**, it must be true that there is an (explicable) *correlation* between **X** and **Y**.

constructions out of simpler representations with non-empty extensions.[35] However, not all empty representations seem to satisfy the model which Dretske proposes for dealing with **PER**. One category of empty representations which I think do not satisfy this model are concepts of would-be natural kinds which are empty through a failure of extension fixation according to the standard model of natural kind semantics (i.e. that of Kripke (1981) and Putnam (1981)).

Lynn Rudder Baker (1991) perhaps has this in mind when she suggests, in a different context, that the view that all uninstantiated properties must be represented, if at all, then by complex representations is implausible. Consider the example of phlogiston, mentioned, without the explicit development which it will be given here, by Baker (1991, p.22). The standard Kripke-Putnam story about the semantics of natural kinds holds that natural kinds apply correctly to a set of items sharing some naturally occurring explanatory property (Boghossian, 1991, p.74). Just what this underlying unificatory property is is determined by a process of extension fixation in relation to local samples (Kripke, 1981, p.122).[36] Although such samples may be identified, and hence the extension of the concept may be *fixed*, by qualitative descriptions (Kripke, 1981) a natural kind concept is not semantically equivalent to (and does not express) such a qualitative description. Indeed the samples can be identified equally well by ostension (Putnam, 1981). For example, the extension of a natural kind term such as 'water' (i.e. the property which it expresses) can be fixed by a definite description such as 'the X such that X is a colourless potable liquid, which fills the lakes and streams on Earth'.

The process of fixing the extension of a natural kind term is mediated by the intentions of the speakers of the language to use the term to pick out a kind of the sort defined above. However, terms which are intended as natural kind terms (i.e. purported natural kind terms)

[35] For example, this seems to account well for our concepts of fictional and mythical entities, since these entities seem to be introduced precisely through this sort of fabrication from other representations (through descriptions, etc.). Concepts of fictional entities and mythical entities, moreover, constitute an important class of empty representations.

[36] That is, the natural kind is that kind exemplified by most of the members occurring in specified local samples (Boghossian, 1991).

can have empty extensions if it turns out that there is no underlying naturally occurring explanatory property exemplified by most of the members in local samples (Boghossian, 1991). This can occur because there is no property or collection of properties which satisfies the description used to fix the extension, or because a large and disparate class satisfies this description. Phlogiston is an example of the former kind: There is no X such that for stuff to burn is for it to give off X (Fodor, 1991).

To return to the objection suggested by Baker. Dretske considers the concept of water (and one may assume other natural kind concepts) to be primitive, since it is syntactically simple and is not semantically equivalent to any description. This follows from the Kripke-Putnam account of the semantics of natural kind concepts.[37] But then the same considerations should surely show that phlogiston is also a primitive concept.[38] It is surely implausible to suppose that as things *actually* are the concept of phlogiston is semantically equivalent to a description, though if phlogiston *had* existed, then it would not be. Just as water never was equivalent to its reference fixing descriptions, so too phlogiston never was and is not now equivalent to its purported reference-fixing descriptions. And, yet, the concept of phlogiston does seem to have content.

Hence, the stipulation that all empty representations must be complex representations will not account for cases of purported natural kind concepts which turn out to have empty extensions.

[37] I am assuming here that this account of natural kind terms in natural language is compatible with a naturalistic account of the semantics of natural kind concepts and thus can be applied (with appropriate alterations) to the latter representations. There seems to be an apparent difficulty with doing this since the standard story about the semantics of natural kind terms makes reference to the intentions of speakers, and the concept of an intention is a semantic notion. However, as Fodor (1990b) notes, we needn't make use of the notion of intention in giving an account of natural kind concepts: What matters is the dispositions which subjects of such concepts have, not the mechanisms by which they come to have such dispositions. Hence, so long as the possessor of a candidate natural kind concept displays the disposition to apply that concept only to members of the kind in question in all such worlds in which such members are distinguishable to the subject from qualitative doppelgangers, then he satisfies the generalization of the condition referring to speaker intention in the case of natural kind terms.

[38] If it is thought to be problematic that phlogiston may be nomologically or metaphysically impossible, and, hence, have no possible world in which it is distinguishable from its qualitative doppelgangers, then we can reframe the example with a purported natural kind which, though actually non-existent, is nomologically possible.

3.2.1.2. Fodor

3.2.1.2.1. The Account

Given Fodor's account of information (**FI**) and the possibility that tokens of a representation type **X** can be caused by not only tokens of some type **Y**, but equally tokens of some type **Z**, with **Z**≠**Y**, some tokens of **X** will carry information about **Y** and some tokens of **X** will carry information about **Z**. However, the content of a representation is common to all tokens of the representation type (Fodor, 1990b). The central question for Fodor, therefore, is 'In virtue of what do **X**s represent **Y** rather than **YorZ**, and hence misrepresent when caused by a **Z**?'.

Fodor's answer to this question is that **X**s represent **Y** rather than **Z** or **YorZ** in virtue of there being an *asymmetric dependence* between **Y**-caused tokens of **X** and **Z**-caused tokens of **X**. Hence, Fodor frames the following *sufficient* (though not necessary) conditions on **X**s expressing the property **Y** (meaning **Y**):

> **(AD)** For any representational state type, **X**, for any property, **Y**, tokens of **X** express (represent) **Y** if:

(i) There is a nomic dependence between the property of being a **Y** and the property of being a cause of **X**s, such that the generalisation '**Y**s cause **X**s' is lawlike.

(ii) For any property **Z**, if **Z**≠**Y** and there is a nomic dependence between the property of being a **Z** and the property of being a cause of **X**s then the law that **Z**s cause **X**s is asymmetrically dependent on the law that **Y**s cause **X**s.

Fodor presents the notion of 'asymmetric dependence' variously. However, the central underlying idea which condition (ii), above, is intended to capture is that that **X**s carry information about **Y**s is the *fundamental* informational relation amongst all of the

informational relations which **X**s bear to the world (Fodor, 1990b, p.91). Letting '**Y**→**X**' represent '**Y**s cause **X**s'[39] we can express this more formally as follows (cf. Fodor, 1990b):

> **(AC)** For any properties **X**, **Y**, **Z**, if **Z**≠**Y** and '**Y**→**X**' is a law and '**Z**→**X**' is a law, then the nomic dependence between **Z** and **X** asymmetrically depends on the nomic dependence between **Y** and **X** if and only if:
>
> (i) If '**Y**→**X**' were to be broken, then '**Z**→**X**' would be broken;
>
> (ii) If '**Z**→**X**' were to be broken then '**Y**→**X**' would still hold.

Fodor prefers to work with the asymmetric dependence condition as framed in terms of nomic dependencies rather than in terms of the interpretation which such nomic dependencies receive in the language of possible world semantics (Fodor, 1990b). Nevertheless he (1990b, p.113) does give a possible worlds gloss on the asymmetric dependence condition. Since, the gloss which he gives does not, to my mind, capture what Fodor intends it to, for present purposes I will work with the following revision of it, based not only on the account which Fodor gives, but equally the use he makes of that account and the amendments he makes to it in response to various objections:

> **(ACP)** For any properties **X**, **Y**, **Z**, if **Z**≠**Y** and '**Y**→**X**' is a law and '**Z**→**X**' is a law, then the nomic dependence between **Z** and **X** asymmetrically depends on the nomic dependence between **Y** and **X** if and only if:
>
> (i) There exists some possible world, W, such that in W '**Y**s cause **X**s' is a law but '**Z**s cause **X**s' is not;
>
> (ii) For every world, W*, if in W* '**Z**s cause **X**s' is a law but '**Y**s cause **X**s' is not, then W is closer to @ than is W*.

[39] This notation is merely a convenience, and should not be interpreted as implying that '**Y**s cause **X**s' entails that **Y**s are nomologically sufficient for **X**s, or that there are non-question-begging conditions which can be specified in which **Y**s are sufficient for **X**s.

3.2.1.2.2. Proposed Solutions to the Problems

HP

On the face of it **(AD)** looks like it will be able to handle **HP** (cf. Fodor, 1990b, p.109). Even though some distal feature, **Y**, on the horizontal axis which causes the tokening of a representational state **X** depends (in some sense) on some more proximal feature, **Z**, causing the tokening of **X**, nevertheless this is not the kind of dependence required for **X** to represent **Y**. For there are a number of distinct proximal routes by which the occurrence of a **Y** can cause the tokening of an **X** in which **Z** tokenings don't feature.[40] Hence, if the **Z→X** connection were to be broken the **Y→X** connection would be left intact. Hence, the latter connection doesn't asymmetrically depend upon the former, and hence **X** means **Y** rather than **Z**.

However, this only shows that in the nearest world in which the **Z→X** connection is broken, the **Y→X** connection is left intact; it doesn't show that this world is nearer to @ than any in which the **Y→X** connection is broken but the **Z→X** connection is left intact. Why should we believe this stronger claim? This question is particularly pressing since it is far from clear that the counterfactual which it is meant to translate, namely 'all else being equal, if the **Y→X** connection were to be broken then the **Z→X** connection would be broken as well' (Fodor, 1990b, p.112), is true. Indeed this counterfactual looks false. Presumably, the reasoning behind this claim is that to get to a world in which the **Z→X** connection is broken but the **Y→X** connection holds, one need only break the **Z→X** connection. However, to get to a world in which the **Z→X** connection holds but the **Y→X** is broken, it would need to be the case that for each distinct proximal 'route' between **Y** and **X** one would need to either break the connection between **Y** and the proximal cause of **X** associated with that route, or one would need to break the connection between the proximal cause of **X** associated with the route and **X**.[41] This, it might be held, would require a greater alteration of the actual world.[42]

[40] This is the familiar starting point which was highlighted in the above discussion of Dretske.

[41] In the present case, considering the route from **Y** to **X** which runs through **Z** one only has the option of breaking the **Y→Z** connection, of course.

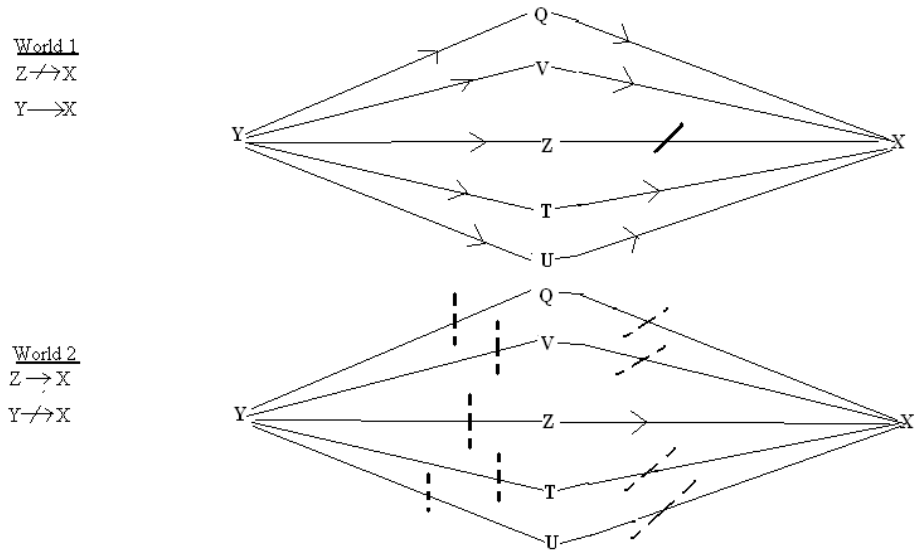

Figure 6

I am not convinced that this move works. However, even if we grant that **(AD)** can eliminate any *one* particular proximal stimulus as the content of **X**, it seems that it cannot eliminate the *disjunction* of all possible proximal stimuli, by which **Y**s cause tokens of **X**, from the content of **X**s. For there are no worlds in which the **Y→X** connection holds but the connection between the disjunction and **X** does not, since it is only by causing the occurrence of one of these stimuli that **Y**s cause **X**s. On the other hand, there are worlds in which the connection between the disjunction and **X** holds and yet in which the **Y→X** connection does not (i.e. those worlds in which the connection between **Y**s and each of these proximal stimuli types does not hold). Hence it seems that the nomic connection between **Y** and **X** is asymmetrically dependent on the nomic connection between this disjunction of proximal stimuli and **X**.

Fodor notes this objection (1990b, pp.109-110) but claims that it is not a problem for **(AD)**. There are at least two lines of argument for this conclusion which can be constructed from his

[42] See figure 6 for a graphical representation of Fodor's response to **HP** as well as the argument in this paragraph. The idea is that world 2 is further away than world one because more nomic connections which hold in the actual world need to be broken to get to world 2. In the diagram a directed line which has been crossed represents a broken nomic connection. The perforated lines represent *possibilities* of breaking nomic connections: The choice, for each possible causal route between **Y** and **X**, in world 2 is between breaking the vertical line or the diagonal line.

work. The first is one which we have already seen above: The disjunction of proximal stimuli is *open* owing to the holistic character of belief/content fixation. However, first, there are surely closed disjunctions either more proximal or more distal from the one which Fodor thinks is open; why doesn't **X** express those disjunctions? More importantly, as has already been discussed above, *in principle*, this line of argument will do Fodor no good.

The second line of argument is not given directly in response to the present problem but rather in a discussion of causal chains (pp. 117-118). Nevertheless it is directly applicable to the case at hand. Fodor's response to the problem of causal chains is to appeal to a further condition on **X**s representing **Y**'s, namely that **X**s be robust with respect to **Y**s, in the sense that it must be possible for non-**Y**s to cause **X**s (so that an **X** occurs without a **Y**). Hence, Fodor might attempt to exclude the disjunction of proximal stimuli as what is expressed by **X**, on the basis that **X** cannot be robust with respect to the disjunction and hence cannot be represented by **X**. Hence, **X**s must express **Y** rather than this disjunction.

However, *prima facie*, we can use the same argument to get to the conclusion that **X** does not represent anything at the same point at which ***Y*** occurs on the horizontal axis. For since **X**s are robust with respect to **Y**s, there are non-**Y**s which can cause **X**s. But, then, consider the disjunction of all these possible non-**Y**s and **Y**. By the above reasoning, **X** is not robust with respect to this disjunction, and, hence, **X**s must represent some still more distal feature than **Y**.[43]

In the light of this objection, the line Fodor would have to take in dealing with **HP** is that **X**s represent **Y** rather than the disjunction of all of the proximal stimuli associated with **Y**s, because **X** is not robust with respect to this disjunction *and* there is no proper subset, $\{\mathbf{P}_j\}_{j\varepsilon[1,m]}$, of the set of the proximal stimuli, $\{\mathbf{P}_i\}_{i\varepsilon[1,n]}$, which is such that all other nomic connections between elements of $\{\{\mathbf{P}_i\}_{i\varepsilon[1,n]}\text{-}\{\mathbf{P}_j\}_{j\varepsilon[1,m]}\}$ and **X** are asymmetrically dependent

[43] Notice that we can iterate this argument for each co-ordinate point on the horizontal axis, to yield the conclusion that **X** has no content at all. In what follows I'll refer to the disjunction of **Y** and all the non-**Y**s featuring at the same point as **Y** on the horizontal axis as 'the disjunction associated with **Y**'.

on the disjunction of elements of $\{\mathbf{P}_j\}_{j\varepsilon[1,m]}$. In particular this is not the case when $\{\mathbf{P}_j\}_{j\varepsilon[1,m]}=\{\mathbf{Z}\}$ (where **Z** as above).

However, on the basis of the reasoning which was put forward in favour of Fodor's being able to avoid the consequence that **X** represents **Y** rather than the specific proximal stimulus **Z** it looks like the closest world in which the nomic relation between the disjunction of elements in $\{\{\mathbf{P}_i\}_{i\varepsilon[1,n]}\text{-}\{\mathbf{Z}\}\}$ and **X** holds but the **Z→X** connection fails is closer than any world in which the **Z→X** connection holds but the nomic relation between the disjunction of elements in $\{\{\mathbf{P}_i\}_{i\varepsilon[1,n]}\text{-}\{\mathbf{Z}\}\}$ and **X** fails. This is because the former world is one in which the nomic relations between each of the elements of $\{\{\mathbf{P}_i\}_{i\varepsilon[1,n]}\text{-}\{\mathbf{Z}\}\}$ and **X** holds but the **Z→X** connection fails, but the latter world is one in which each of the nomic relations between the elements of $\{\{\mathbf{P}_i\}_{i\varepsilon[1,n]}\text{-}\{\mathbf{Z}\}\}$ and **X** fails but the **Z→X** connection holds. Hence we have to change substantially more about the actual world to get to the latter world than we do to get to the former world. Consequently, since $\{\{\mathbf{P}_i\}_{i\varepsilon[1,n]}\text{-}\{\mathbf{Z}\}\}$ is a proper subset of $\{\mathbf{P}_i\}_{i\varepsilon[1,n]}$, let $\{\mathbf{P}_j\}_{j\varepsilon[1,m]}$ be the singleton constituted by the disjunction of all of the elements in $\{\{\mathbf{P}_i\}_{i\varepsilon[1,n]}\text{-}\{\mathbf{Z}\}\}$. Then this is a proper subset of the set of the proximal stimuli which is such that all other nomic connections between the remaining proximal stimuli and **X** are asymmetrically dependent on the disjunction of elements in this subset. Hence, this disjunction cannot be excluded from being a contender for constituting the content of **X**.

Of course, as I noted above, I don't find the reasoning which was put forward in favour of Fodor's being able to avoid the consequence that **X** represents **Y** rather than the specific proximal stimulus **Z** entirely convincing. However, the present point is just that *if* that line of reasoning is cogent then Fodor can avoid the consequence that **X** represents some specific proximal stimulus **Z**, but he can't avoid the consequence that **X** represents some disjunction of such possible stimuli. On the other hand *if* the line of reasoning is faulty then it looks like Fodor can avoid the consequence that **X** represents some disjunction of such possible stimuli, but he can't avoid the consequence that **X** represents some specific proximal stimulus **Z**. Either way there's trouble for Fodor.

Hence, I conclude that **(AD)** doesn't adequately solve **HP**.

VP

(AD) also does not adequately solve **VP**. Since **DP** is a special case of **VP**, and since **(AD)** is designed primarily to deal with **DP** (Fodor, 1987, 1990b) its ability to adequately address **VP** is really the litmus test of the proposal.

In terms of one aspect of **VP** Fodor's response is the same as Dretske's: A representational state **X** has as its extension **Y**s rather than some subset of **Y**s or some superset into which all **Y**s fall because the nomic connection is between the property of being a **Y** and the property of being a cause of **X**s rather than the property which defines the sub-/super-sets of the extension of **Y** and **X**.

As has been gestured toward in the previous subsection the problem of **DP** is addressed by Fodor (1990b) as follows. In any case in which **Y→X** is a law and **Z→X** is a law, it is also true that **YorZ→X** is a law. However:

(i) **X**s will express **Y** if the nomic connection '**Z→X**' depends asymmetrically on the nomic connection '**Y→X**';

(ii) **X**s will express **YorZ** if the nomic connections '**Z→X**' and '**Y→X**' depend symmetrically on one another;

(iii) **X**s will express an ambiguous meaning between **Y** and **Z** if the nomic connection '**Z→X**' is independent of the nomic connection '**Y→X**'.

(AD) is forwarded as providing sufficient conditions for representational content, as are conditions (i), (ii) and (iii), above. Hence, Fodor claims that were the nomic dependencies as described in (i), then **X** would have a non-disjunctive content. Moreover, he claims that since he is only offering *sufficient* conditions the analysis can be undermined only by developing

cases in which the purportedly sufficient conditions obtain and yet in which the content of the state is not as the theory claims. Hence, Fodor seems to regard as irrelevant arguments to the effect that **(AD)** and its purported solution to **DP** are incorrect on the basis that it seems implausible to think that in the case of our paradigmatic examples of mental content Fodor's conditions are satisfied (cf. Fodor, 1990b).

Now I should note at the outset that I doubt that Fodor is right in adopting this line, for much the same reasons as were mentioned in section I, above. It seems to me that the primary (only?) way of assessing whether a theory which purports to give sufficient conditions for meaning is really giving an account of *that* phenomenon is to test such conditions against our paradigmatic examples of meaning/content. When we do that however, it seems that **(AD)** is a non-starter. To see this consider a characteristic example of error (taken from Fodor (1987)): The mistaking of a slim cow seen at a distance for a horse. Presumably in this case the error is explainable by the fact that the cow looked, at that distance, like a horse, that it cast the same pattern of light onto the retina as a horse would have in those circumstances etc.. But then it seems that if these are the sorts of cases which establish the cow→**HORSE** nomic connection, then the easiest way in which this connection could be broken is to sever the causal connection between these sorts of retinal projections and tokens of **HORSE**.[44] This is presumably easier than severing the connection between slim cows seen from a distance and those retinal projections, since in this case reasonably large changes in bovine genetics or terrestrial optics would be required.[45] Given this, however, it looks like the nearest worlds in which the cow→**HORSE** connection is broken are one and the same as the nearest worlds in which the horse→**HORSE** connection is broken. Hence, there really is no asymmetric dependence here.

[44] It should be emphasised that '**HORSE**' is being individuated intrinsically. Hence, all that is required to sever the *retinal projection*→**HORSE** connection is for the causal 'wiring' mediating the connection to be slightly different.

[45] It should be emphasized that we can't sever the *nomic* connection between cows and **HORSE**s by simply ensuring that there are no slim cows seen from a distance. For in this case it is still true that *were* a slim cow to be seen from a distance then it *would* cause a tokening of **HORSE**, and, as Fodor says, it's the counterfactuals that count in semantics.

Of course it is likely to be objected to this that this description of the breaking of the cow→**HORSE** connection, does not entail that were this connection to be broken then so would the horse→**HORSE** connection, since one can identify a horse through more than one sense modality. However, though it is surely *possible* to do this, it is equally possible that someone who is *unable* to do this can nevertheless represent horses and would not be confined to merely disjunctive content with respect to these animals. It seems to me that this sort of example generalises, and, hence, casts doubt upon the idea that we have any good reason to think that **(AD)** really is an account of meaning and hence that it really does solve **DP**.

Nevertheless, I will not rest my case on this line of argument. It is far less obvious that cases of natural kind terms can be included in the generalisation gestured toward in the previous paragraph, and it will be more satisfying if we can undermine **(AD)** on its own terms. I will therefore forward two lines of argument in what follows, both of which take **(AD)** as its maker intended.

The first argument is a development of one given by Godfrey-Smith (1989, pp.538-540). He argues that the lesson of **DP** is that it is the nomic link between the disjunctive kind **YorZ** and **X** which is the fundamental one. Hence, the nomic connections between **Y** and **X** and between **Z** and **X** are two *instances* of this fundamental connection. As such, it would seem that, in this case, the question of asymmetric dependence just simply does not arise unless one, *question-beggingly*, assumes that the representational content of **X** cannot be **YorZ**.

I think that Godfrey-Smith's argument is a good one. However, one might question whether there is only one nomic connection here. Fodor is a realist about properties and nomological relations (1990b), and hence would take there to be a fact of the matter about whether there are three nomic relations here or just one. Although Godfrey-Smith is right that Fodor cannot simply assume that there are three distinct connections, by the same token we cannot simply assume that there is only one.

However, suppose that there really are three independent nomic connections in the case at hand.[46] On this assumption '**YorZ**' expresses a distinct property from either '**Y**' or '**Z**'. Hence, we need to decide which nomic connection (i.e. that involving **Z**, **Y** or **YorZ**) is the asymmetric base.[47] But now it seems clear that this base must be the **YorZ→X** connection, since it is true that if **YorZ**s ceased to cause **X**s, then so would **Y**s and **Z**s, though not vice versa. [48] Hence, **X**s must have a disjunctive content.

Hence, on the first line of argument there is only one nomic dependence which obtains between **YorZ**s and **X**s, and hence **X**s cannot represent **Y**s but can only represent the disjunctive content **YorZ**. On the second line of argument, though **X**s can express **Y**, **Z**, or **YorZ**, by **(AD)** they must express the last of these, and hence **X**s have a disjunctive content. Either way, therefore, **DP** has not been solved.

In general, Fodor's defence of the claim that it is *possible* for his conditions for content to be met depends on the premise that *ceteris paribus* the more things one has to change to transform the actual world into a distinct possible world, the harder it is to get to the latter world (1991, p.277).[49] Moreover, demonstrating the possibility of a naturalistic semantics does require such a defence, even if one is only offering sufficient conditions. However, given this principle, it is always possible to construct a deflationary disjunction which will have equal claim on **(AD)** as some non-disjunctive item for the title of 'proper content' of some representational state.

---

[46] I am assuming here, what is generally assumed by all parties in the debate, that '**YorZ**s cause **X**s' is counterfactual supporting.

[47] I borrow this term from Boghossian (1991) to refer to that nomic connection on which all others asymmetrically depend.

[48] **DP** gets off the ground only by assuming that for every case in which there are laws '**Y→X**' and '**Z→X**', there is a property **YorZ**, such that '**YorZ→X**' is a law and if a causal interaction is covered by the law **Y→X**, then it is covered by the law **YorZ→X**, and if the causal interaction is covered by the law **Z→X**, then it is covered by **YorZ→X**. Whence the truth of the statement in the text.

[49] See, e.g. his (1990b) discussion of the water/XYZ case.

To see this, let **Y→X** be a law and the purported asymmetric dependence base. Suppose further that the set of possible non-**Y**s which can cause **X**s is $\{\{\mathbf{Z}\}U\{\mathbf{Q}_i\}_{i\varepsilon[1,n]}\}$. Now, consider the disjunction of all of the members of the set $\{\{\mathbf{Y}\}U\{\mathbf{Q}_i\}_{i\varepsilon[1,n]}\}$. The nearest **Y→X** world in which **Z→X** fails, W1, is a world where $\mathbf{YorQ_1or...Q_n}\rightarrow\mathbf{X}$ holds but **Z→X** fails (since '**Y**' is in the disjunction). It is, moreover, the nearest such world, since $\mathbf{Q_i}\rightarrow\mathbf{X}$ must hold in W1 for all iε[1,n], in order for W1 to be as close a **Y→X** world to @ as possible. However, the *nearest* **Z→X** world in which **Y→X** fails, W2, would be a world in which $\mathbf{Q_i}\rightarrow\mathbf{X}$ holds for all iε[1,n]. Hence, W2 is also a $\mathbf{YorQ_1or...Q_n}\rightarrow\mathbf{X}$ world. But then the nearest world in which $\mathbf{YorQ_1or...Q_n}\rightarrow\mathbf{X}$ fails but **Z→X** holds must be further away from @ than W2, and *de facto* further away from @ than W1.[50] So, $\mathbf{YorQ_1or...Q_n}$ is the asymmetric dependence base. Moreover, **X** is robust with respect to it.

Hence, if Fodor's line of reasoning in demonstrating that it is possible for the conditions in **(AD)** to be satisfied, then it turns out that he hasn't solved the disjunction problem.[51] On the other hand if his reasoning is not sound, or if there is something wrong with the distance measure which he uses then he has not demonstrated the possibility of the conditions in **(AD)** being satisfied. Either way, and taking the account on its own terms, **(AD)** is unsuccessful in addressing **DP**.

There is a common intuitive thread tying these previous two arguments together with the central argument against Fodor's proposed solution to **HP**. Fodor seems to assume that he can articulate conditions for when a representation will count as disjunctive and when it won't simply in terms of those expressions which pick out the relevant non-disjunctive contents. However, once we apply the asymmetric dependence condition directly to the properties picked out by well-chosen disjunctive predicates things fall apart. Hence, if **(AD)** is to solve **HP**, **VP** and **DP**, it must prioritise the non-disjunctive predicates over the

[50] Note that there is no world between W1 and W2 in which $\mathbf{YorQ_1or\ldots orQ_n}\rightarrow\mathbf{X}$ fails but **Z→X** holds. For if there was, this would then be a world in which **Y→X** fails but **Z→X** holds which is nearer to @ than W2. But W2 is the nearest such world. Contradiction.

[51] The argument requires that there be three or more possible distinct types of item which can cause the tokening of a representational state. It seems plausible that this assumption is usually satisfied, however.

disjunctive ones. But that is just what a consideration of these problems suggests one cannot do.

PNC

The approach taken by Fodor to **PNC** is essentially the same as that of Dretske (1981), discussed above. Hence, I will give it no further discussion here.

PIR

There are two distinct approaches taken by Fodor to **PIR**. The first of these is very similar to Dretske (1981).[52] The second is the, to my mind much more plausible and more promising, approach given in Fodor (1994). He argues there that we can differentiate the contents in **PIR** by amendment of a pure informational semantics with an appeal to the *inferential role* which the representations play in the cognitive economy of the subject. He demonstrates this approach by using the following example, based on figure 7, below (taken from Gates (1996, p.341)).

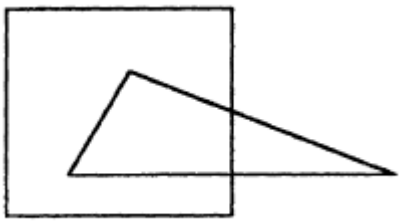

Figure 7

[52] One point of difference (which doesn't make a difference to the success of the account) is pointed out by Gates (1996) who claims that Fodor's view differs from that of Dretske in that whilst the latter takes the semantic contents of **RABBIT**, **UNDETACHED RABBIT PART** and **TEMPORAL RABBIT STAGE** to be the same, Fodor, takes each to be different. That is, only a complex representational state can express *undetached rabbit part* or *temporal rabbit stage*. I agree with Gates (1996) that this is a mistake.

Fodor considers the case in which a subject, S, has two representational terms **SQUARE** and **TRIANGLE**, which are caused to be tokened by instantiations of *square* and *triangle* respectively. Since the properties of being a square and being an undetached part of a square are necessarily co-instantiated, as are the properties of being a triangle and being an undetached part of a triangle, the question arises as to whether **SQUARE** means *square* or *undetached proper part of a square*, and similarly for **TRIANGLE**.

Fodor's (1994, pp.64-74) solution to this problem is to consider the following two hypotheses: (a) that **SQUARE** means *square* and **TRIANGLE** means *triangle*; (b) that **SQUARE** means *undetached proper part of a square* and **TRIANGLE** means *undetached proper part of a triangle*. If (b) is true then there will be circumstances under which S tokens the mentalese sentence 'A is a square', tokens the sentence 'A is a triangle', infers from this that 'A is a square and A is a triangle' and infers from *this* that 'A is a square and a triangle'.[53] On the other hand, if (a) is true then there would be no circumstances under which this pattern of inference will occur. Hence, there are circumstances under which we can differentiate these two hypotheses and reject one in favour of the other.

I think that the move to examining the internal relations which representational states bear to one another in addressing **PIR** is a good one. However, I don't think that Fodor's particular treatment of the problem is successful. Why should we accept the premise that (i) if S's tokens of **SQUARE** mean *square* and her tokens of **TRIANGLE** mean *triangle* then there are no circumstances under which she will accept the above pattern of inference, but (ii) if S's tokens of **SQUARE** mean *undetached proper part of a square* and her tokens of **TRIANGLE** mean *undetached proper part of a triangle* then there *are* circumstances under which S would accept the inferences? The reason is surely that it is *impossible* for anything to be both a square and a triangle simultaneously whereas it *is* possible for something to be an

[53] This final inference determines that A refers unambiguously.

undetached proper part of a triangle and an undetached proper part of a square simultaneously. But why should that constrain the inferences that S is prepared to accept?[54]

The point is this: Even when a subject possesses two representational terms expressing ***X*** and ***Y***, it needn't follow that S grasps all of (perhaps any of) the metaphysical necessities which obtain between **X**s and **Y**s.[55] Hence, it is implausible to think that *in general* one can only possess representations which express certain properties if one's pattern of inference involving those representations reflects the metaphysical necessities which obtain between these properties. Fodor would need to give us reasons as to why the **PIR** case is special in this regard, which he doesn't.[56]

There is also a second problem with his strategy, however. In virtue of what would the distinct patterns of inference determine the extensions of the terms involved? In virtue of what, that is, would patterns of inference determine that the terms involved expressed different properties in the world, rather than expressed the same property, thought about in two different ways? Plausibly the concepts of water and H_2O have different inferential roles without it following that the properties expressed by these concepts are one and the same.

[54] Remember, although we are picking out the relevant properties by the descriptions 'square' and 'undetached proper part of a square' it is that which is picked out by the descriptions which is expressed by S's representational terms, rather than the descriptions themselves. We are also assuming that it is not the case that in having a representational primitive which expresses *square* and another one which expresses *triangle* S grasps some description of the properties of squares and triangles from which it follows that nothing can be a square and a triangle simultaneously. At least, if we are assuming this then the example of squares and triangles is inappropriate to the case of rabbits and undetached rabbit parts.

[55] Indeed this is one of the lessons we can take from **PNC**.

[56] Could Fodor claim that this only shows that a representation having an inferential role which mirrors the metaphysical necessities of some item about which the representation carries information is not *necessary* for the representation expressing that item, and hence change his claim to one about sufficiency? Not obviously. For so long as he allows that a representation's inferential role may not reflect the metaphysical necessities which its represented bears to other items in the world, it is possible for a state express *triangle*, whilst having an inferential role which: (i) mirrors the necessities which *triangle* bears to other properties, and, (ii) does *not* mirror the necessities which *undetached proper part of a triangle* bears to other properties. But if the mirroring between inferential role and metaphysically necessary relations which the property bears to other items is not necessary for content, then the possibility is not excluded that this representation expresses *undetached proper part of a triangle* as well. Hence, this move does not necessarily solve **PIR**.

Again Fodor would need to give principled reasons for thinking that the **PIR** case is distinct here, which he doesn't.

What we need in order to make Fodor's suggestion work in the light of these two points is some principled reason to think that the external, meaning co-determining relations which the representation bears to some item in the world somehow constrain the causal patterns that the representation can enter into. However, the relations appealed to in a pure informational semantics cannot provide the basis for such a reason.

Finally, Fodor's consideration of what inferences a subject would accept threatens to render the account of content holistic. This is because which inferences one is prepared to accept depends upon which other background beliefs one has (Sterelny, 1990).[57] Of course not all will find this consequence problematic and demonstrating that it *is* genuinely problematic would take us beyond the bounds of the present paper. Hence, I will just note that being committed to a holistic theory in the present context is a rather unwelcome consequence, for two reasons. First, Fodor explicitly formulates the project of informational approaches to semantics as that of providing an *atomistic* account of representational properties. Second, it is intuitively difficult to see how any content *could* be fixed on a holistic account, since what content any one state has depends upon the content of a large portion of one's other states, the content of any one of which depends in turn on the content of still other states, etc.. Either this requires the attribution of infinitely many representational states or it requires that some states both have their content determined by, and simultaneously determine, the content of other states. Neither option is entirely palatable.

PER

Prima facie the most promising application of **(AD)** to the first five problems of determinacy of mental content is its application to **PER**. As it stands **(AD)** does not require there to be any

[57] The degree of holism is not confined since we can construct Quine-type cases for very many distinct representational types.

actual causal interaction between the property expressed by a representational state and that state. So long as the property features in an appropriate nomic relation which satisfies the conditions for being the asymmetric dependence base, the property can be represented.

However, this promise is undermined by the following point. Baker (1991) notes that, for any empty representation, we can construct a qualitative 'doppelganger' of the property expressed. For example, the property *unicorn* has the 'doppelganger' *shunicorn*[58] where shunicorns are non-unicorn, white, horse-like, horned creatures.[59] Since, both unicorns and shunicorns are so similar, it seems plausible to think that shunicorns would cause **UNICORN** tokens if unicorns did. Moreover, their similarity also suggests that the nearest world in which the *unicorn*→**UNICORN** connection held but the *shunicorn*→**UNICORN** connection failed would be the same distance from the actual world as the nearest world in which the *shunicorn*→**UNICORN** connection held but the *unicorn*→**UNICORN** connection failed. Hence, **UNICORN** would have a disjunctive content. Since we can construct such doppelgangers for any empty representation, it seems that Fodor's account entails that every empty representation will have a disjunctive content, which seems incorrect.

The central point (to which Baker (1991) also alludes) is that Fodor's theory is one which depends essentially on relations between laws across counterfactual scenarios. However, in the case of uninstantiated properties it just isn't clear what *would be* required for them to be instantiated. Hence, it just isn't clear what the relative distances of worlds from @ are, and, consequently, whether it *is* possible for Fodor's conditions to be satisfied in such cases.[60]

[58] The term, though not the description, is due to Baker (1991).

[59] We can assume that shunicorns differ from unicorns in some relevant way (perhaps genetically), or in certain qualitative features which are detectible only with some scrutiny.

[60] Fodor (1991) does respond to this argument. However, I find his response unconvincing. He claims that since he is only offering sufficient conditions for content, he gets to stipulate which nomic relations hold. The burden on the critic's shoulders is then to construct cases in which the proffered conditions are satisfied and yet the meaning isn't as predicted. Baker's objection clearly doesn't have this form, since it seems to be questioning whether the sufficient conditions for content have been satisfied in the unicorn/shunicorn case. Hence, Fodor implies that he needn't engage with Baker's line of argument.

However, this underestimates the force of Baker's objection. The task which Fodor has taken on is that of showing how it is *possible* for there to be representational content in a naturalistic world, and, *de facto*, how it is

Undoubtedly Fodor has got something right in the idea of asymmetric dependence, for the relation between a representation and what a representation represents is, in some sense, more fundamental that the relation between the representation and anything which it misrepresents. Moreover that property which is expressed by a representation is in some sense responsible for the representation in a way which properties misrepresented by the token are not. However, the above discussion (particularly the discussion of **HP**, **VP**, and **DP**) seems to reveal that this asymmetry is a *consequence* of some other meaning related condition rather than the foundation of all such conditions.

---

possible for there to be empty representations. In order to do this it won't be enough to simply give a set of purportedly sufficient, naturalistic conditions and then stipulate that they hold. Rather one must demonstrate that it is *possible* for them to hold. This is what Fodor attempts to do for the water/XYZ case, by attempting to demonstrate the existence (and hence possibility) of some mechanism which would mediate the asymmetry of the dependence. However, Baker's objection highlights: (i) the lack of demonstration on Fodor's part that it is possible for the conditions of **(AD)** to be satisfied in the case of empty representations; (ii) the lack of any principled basis on which this demonstration might proceed (cf. Baker, 1991, p.20).

In particular, Fodor's specific response to the water/XYZ case cannot work for the unicorn/shunicorn case. Since neither unicorns nor shunicorns are instantiated, we cannot apply the standard Kripke/Putnam account of natural kind terms to unicorn representations. Hence, it does not seem that there are differential dispositions with respect to the tokening of a representational state in the presence of unicorns as opposed to shunicorns.

Hence, Baker's objection attempts (and succeeds in my view) to give us reason to believe that the conditions in **(AD)** *could* not be satisfied in the case of empty representations.

IV

Teleological accounts

In order to solve the determinacy problems whilst avoiding the objections raised above we need a further source of discrimination between allowable contents (Sterelny, 1990). In search of this some philosophers have made an appeal to biological norms governing the functioning of the *producers* or *consumers* of internal representational states. In what follows I will consider one prominent example of each type of approach.

4.1. Teleological Accounts Focusing on the Representation Producer

Dretske (1986, 1988) amends his informational semantics by introducing a teleological element. As has been seen above in his (1986, 1988) account of representational content, Dretske adopts the weaker notion of information, $\mathbf{DI_2}$; as has also been seen (and as Dretske (1988) notes), however, on neither version of information/indication is error possible. Hence information/indication cannot be the representation relation.

As such, Dretske (1986) provides the following teleological version of indicator semantics:

> (**TIS**) An inner state type **X** represents a type of state of affairs, **Y**, iff **X** has the function of indicating **Y**.

4.1.1. Proper functions, selected effects and the selection of/selection for distinction

Although Dretske (1988) is relatively non-committal about the relevant notion of function in **(TIS)**, we can construe this notion as that proposed by Wright (1973, in Neander, 2004): The function of an item is an effect of the item which "account[s] for it being there" (Wright 1973, p. 364, in Godfrey-Smith, 1992). As Dretske (1988) implicitly holds, this notion of

function can be realised through learning or natural selection.[61] In the case of the latter, the notion becomes that of biological proper function: For any item, **p**, the phenotypic expression of the genotype, **g**, **p**'s proper function is **q** iff that **p** caused **q** causally contributed to the selection by natural selection of **g** (Neander, 2004).

Hence proper functions are effects that have been selected *for*. Here, for some phenotypic expression, **p**, of the genotype, **g**, there is selection *for* **p** just in case **p** causes the selection of **g** by natural selection; there is selection *of* **p** just in case **p** is selected by natural selection (Sober, 2010). Hence, the phenotypic property selected *for* is that property which is causally explanatorily relevant to the fact that the genotype responsible for it was selected.

This explanatory relevance is revealed by the consideration of appropriate counterfactuals (see Fodor, 2008; cf. Sober, 2010). Roughly, these counterfactuals would be of the form 'were **g** not to have had phenotypic expression **p**, then there would not have been selection of **g**' and 'were a relevantly similar genotype, **g***, to have phenotypic expression **p**, there would be selection of **g***' (cf. Horgan, 1989).

4.1.2. Proposed Solutions to the Determinacy Problems

The central benefit of an appeal to teleology is that it allows the information theorist to grant that states can indicate more or less reliably (within bounds) and can indicate more than one type of state of affairs, whilst still having the *function* of indicating one rather than another of these types. An internal state **X** accurately represents when it is caused by states the indication of which is selectively responsible for the existence of the state; and it misrepresents just in case it is caused by a state of affairs which it is not its function to indicate.

[61] Dretske holds that it is only the former which can do the job required of function ascriptions in the case of accounting for belief contents. However, I agree with Godfrey-Smith (1992) that this rests on a mistake. Hence, I will ignore this qualification here.

HP

In terms of **HP**, *prima facie* an appeal to teleology can provide assistance. Although a state **X** may provide perfectly reliable indication only for *proximal* causes (or for no causes at all), it may still represent distal causes if it is its function to indicate such states of affairs. Although there may be more than one 'horizontal' type of state of affairs which is indicated, a state represents only that state which it is its function to indicate. Hence, *on the face of it* an appeal to teleology can avoid the problems of distality which were raised for $\mathbf{MC_{D1}}$.

This face value in addressing **HP** is a little misleading however, a point Dretske (1986) notes via his 'anaerobic bacteria' example. Certain northern hemisphere anaerobic bacteria have tiny magnets called magnetosomes which cause the bacteria to move downwards away from the oxygen-rich, and, therefore, toxic, surface water. It would seem that the proper function of these magnetosomes is to indicate the location of water with a low concentration of oxygen, since this seems to be the feature of these structures which contributed to the survival and proliferation of the species. However, if this is the proper function of the magnetosomes then when they direct the bacteria into oxygen rich surface water owing to the reversal of the direction of local magnetic north (say because a magnet is held above the bacterium) we must conclude that the magnetosome is malfunctioning.

Intuitively, however, this seems incorrect. It is possible for the magnetosome to direct the bacterium into oxygen depleted water *only by* indicating the direction of local magnetic north. Hence, intuitively, it would seem that it is not the magnetosome which is functioning in a biologically abnormal way, but rather the *environment* which the bacterium finds itself in which is biologically abnormal relative to the bacterium. At least it is not clear what basis biology gives us for choosing between these two descriptions.

Of course, as Dretske notes, this sort of example can be seen to generalize. The approach to **HP** adopted in an account such as **(TIS)** is to appeal to biological considerations in order to exclude specific proximal stimuli as the content of a state, **X**, given that it is some distal stimulus which is explanatorily relevant to the selection of the state type (or the mechanism

which produces it). However, given that **X**s can only indicate the distal stimulus by indicating some *disjunction* of more proximal stimuli, in any world in which there is selection of the indication of the distal stimulus by **X**, there must also be selection of the indication of the disjunction of proximal stimuli by **X**. Hence in the actual world if there is selection *for* the indication of the distal stimulus there must be selection *for* the indication of some appropriate disjunction of proximal stimuli. [62] [63]

In parallel with the discussion in section III, one might require that this disjunction include proximal stimuli that cause tokenings of **X** only in counterfactual scenarios. Hence, one might wonder whether such a disjunction would not be infinite and whether this would not cause a problem for the notion that the indication of this disjunction is selected for. Indeed Dretske's (1986) own response is to require that the organism, in order to have determinate distal content, must be capable of basic associative learning. By the argument considered in section III, Dretske believes that this will solve the problem of distality since it will render the class of proximal stimuli as infinitely large.

This line of argument has been considered in detail in the previous subsection, and the same points raised there apply in the present context as well. In particular, if this argument is to do the work which Dretske requires of it, then he must claim that *at the time of selection*, because there were infinitely many proximal stimuli which, through associative learning, *could* come to trigger tokens of the internal state, **X**, there could only have been selection *for* the indication of the distal stimulus, **Y**. As such, in the context of selection, the mechanisms of associative learning must be allowed by the background conditions which purportedly ensure the non-accidental correlation between **X** and **Y**. However, if this is so, then in that context it would be possible to generate infinitely many secondary conditioned stimuli (at

[62] This problem also poses difficulties for an account of misrepresentation, for it would seem that there would always be a deflationary redescription (Dennett, 1987) of the proper function of the representational state which renders misrepresentation impossible.

[63] I take this problem to be a special instance of the general problem of complex causal roles. In what follows, I will show that Dretske's particular solution to it is unsuccessful. In section V the general problem and proposed solutions to it will be discussed in detail.

least one for each primary conditioned stimulus). Since we can ensure that these secondary conditioned stimuli do not indicate **Y**, it would be possible to ensure that **X**s don't indicate **Y**, in the context of selection. If one opens the door to allow the mechanisms of associative learning in, it is impossible to prevent the indication of the distal stimulus by the representational state from slipping out.[64]

VP

In terms of **VP** an appeal to teleology can, *prima facie*, assist us in solving the disjunction problem. The disjunction problem arises from the inference from (i) a state **X** misrepresents by having content ***Y*** but being caused by a **Z** (≠**Y**), to (ii) **X** has content ***YorZ*** and, thus, doesn't misrepresent when caused by a **Z**. But the proponent of **TIS** needn't accept this inference. On **TIS X** represents **Y** iff the function of **X** is to indicate **Y**. Now, so long as in the case of a token **Z** causing a tokening of an **X**, the correlation between **X** and **Z** is not causally efficacious in **X**'s selection, then that **X** indicates **YorZ** adds nothing over and above the fact that **X** indicates **Y** to the explanation of the selection of **X**.[65] Hence, the proper function of **X** is to indicate **Y** rather than **YorZ**. Thus, **X** represents **Y** rather than **YorZ,** despite being sensitive to/correlated with **Z**s.

This is a promising move, but offers only a partial solution to **DP**, since there are cases in which (intuitively) a representational state, **X**, misrepresents by being caused by some **Z**, the indication of which is causally efficacious in the selection of **X**. These sorts of cases arise in the context of the problem of complex causal roles, **PCR**. Hence, the ultimate success of **(TIS)** in solving **VP** and **DP** is contingent upon a satisfactory solution to **PCR**. In section V and section VI, however, it will be argued that the most promising approach to **PCR** requires an abandonment of a producer-based teleosemantics of the form of **(TIS)**.

[64] Of course secondary conditioned stimuli are in a sense causally explanatorily irrelevant to the process of selection, because the benefit of producing a certain behavior, *by hypothesis* is secured only by the presence of the distal stimulus. The current point, however, is not that a disjunction of proximal stimuli including secondary conditioned stimuli would be selected for; it wouldn't. The point is rather that if we are allowed to consider those proximal stimuli which could come to trigger the internal state through a process of associative learning, it can no longer be the *indication of the distal stimulus* which is selected for.

[65] Alternatively, though not equivalently (see n.9), as long as the fact that **X**s indicate **YorZ** is not causally efficacious in **X**'s selection, then that **X** indicates **YorZ** is not selected for.

PIR

In the case of **PIR**, the relevant properties are necessarily co-instantiated. Hence, any internal state carrying information about the one carries information about the rest. Now in principle this latter point is not a problem for **TIS** since, as has been noted above, the account can differentiate between two states of affairs types indicated with the same reliability. However, the former point is a problem, because it entails that in any possible world in which **X** carries information about, or indicates, one Quine-type property, it carries information about, or indicates, the rest as well. Hence, the indication of one of these properties cannot be selected for without the indication of each of the other properties being selected for. Hence, **X** cannot have the function of indicating one Quine-type property rather than another, and hence cannot have a content expressing one of these properties rather than another.[66]

PNC and PER

Finally, with regard to **PNC** and **PER** this kind of appeal to teleology, on its own, seems to offer no further assistance. Teleological notions are appealed to in order to provide further discriminative machinery in determining which out of a range of properties is expressed by a mental representation. They promise to do this by appealing to certain aspects of the history of the state type. However, the property expressed by each of two necessarily co-extensional states is the *same* property, and that which is expressed by an empty representation never played any role in the history of the state type.

This being said I do think that there is some value in an appeal to teleology in responding to **PNC** and **PIR**. However, I think that this value can only be appreciated once we take into consideration properties of the representational system occurring at the level of the system's *internal functioning* rather than simply at the level of the informational relations which it bears to the world. This point will be explored in the following subsection.

[66] Even if we could separate out the indication of these properties, we would still have to show that there is a selective *advantage* to indicating the one rather than the other, and it might be doubted whether there is such an advantage (see, e.g., Loewer, 2002).

Even the partial success of **(TIS)** is undercut by a fundamental problem with grafting a teleological element onto an informational/indicator semantics. Godfrey-Smith (1989, 1992) has argued persuasively for the claim that the explanatory role attributed by Dretske to the indication relation is misplaced. He suggests that considerations concerning the costs and benefits of different *kinds* of mistakes which organisms can make (particularly those of producing false positives and false negatives, respectively) are central to the explanation of organism-world relationships. This often displaces the explanatory role of indication without thereby displacing our intuitive attributions of representational status. That is indication is not necessary for representation. Similar ideas are present in Millikan (1989) and Stich (1990).[67]

As an example, consider a predator-prey situation of the kind described by Millikan (1989, p.288). In a population of beavers, in the presence of danger, one beaver slaps its tail on the surface of the water. The consequence of this splashing is to cause all of the other beavers to find protective cover in a location corresponding to physical features of the tail splashing. Now, supposing that it makes sense to talk about representation in such a context, it would seem to be plausible that this splashing system has the function of representing danger to the other beavers, such that predatory danger is represented by certain physical features of the splashing to be at a certain location relative to the splashing beaver.

However, considerations of costs and benefits to the beavers strongly suggest that, since this system functions to cause the beavers to find protective cover, it is entirely possible that given the constraints on the resources available to the beavers, the process of natural selection would favour the production of false positives (i.e. signalling danger when in fact there is none) rather than false negatives. Signalling danger when there is none imposes relatively limited costs on the beavers, but not signalling danger when there is such danger imposes the ultimate cost on the beavers, i.e. their death.

[67] This argument runs also against Lloyd's (1987) weaker account of information requiring only that the conditional probability of the represented condition on the occurrence of the internal state exceed the conditional probability of any other relevant alternative condition on that state's occurrence (Godfrey-Smith (1992)).

Hence a system which tends to produce false positives would, in certain circumstances, be favoured by evolutionary selection over a system which produced false negatives. Equally, just a few occasions on which such a system signalled danger when in fact there was such, may be enough of a contribution to the overall fitness of the beaver species, for that system to be selected for such signalling (Millikan, 1989). Still, since we initially took the system to produce *representations* of danger, and since the system may have been selected *not* for indicating danger, but rather for being 'overly sensitive' to the 'evidence' for danger, we must hold, it seems, that a structure can be a representation yet not indicate, not have indicated, and, *de facto*, not have the function of indicating, that which is represented.

In general, therefore, contexts in which the relative costs (including those relative to the resources available) and benefits to a species result in the processes of evolution and natural selection favouring systems which produce false positives rather than false negatives will generate counterexamples to **(TIS)**. This, so long as we take that account to be offering both *necessary* and sufficient conditions for a representational state to have a certain content (Godfrey-Smith, 1992).[68]

4.2. Teleological Theories Focusing on the Representation Consumer

One way in which to avoid this last problem is by turning to the representation consumer. For Millikan (1984, 1989) both the content of a representation, and the status of an internal state *as* a representation, is dependent upon its *relation* to the functioning of the representation

[68] Might we then simply claim that **(TIS)** provides sufficient, though not necessary, conditions for representation? Not easily. The above line of argument suggests that once we take biological/evolutionary considerations into account, we see that representation can get on perfectly well *without* indication. Hence, the burden then falls on the proponent of an account such as **(TIS)** (construed now as only providing sufficient conditions for representation) to provide good reason for thinking that, within a biological framework, in those cases in which states *do* indicate that which is represented, the indication relation plays a *non-redundant* part in a set of sufficient conditions for representational content. However, it is not at all clear that this is true. Once we account for the representation relation in those cases in which there is *no* indication, it seems plausible that the combination of such an account with a consideration of the cost-benefit considerations obtaining in cases of representation *with* indication will displace the notion of indication within the explanatory framework offered by the account. That is, it seems plausible that indication may turn out to be a by-product of, rather than a foundation for, representation.

consumer (i.e. the system which *uses* the representation), rather than the representation *producer*. For Millikan (1984, p.107; 1989, pp. 286-287):

> (**TRP)** An inner state **X** is a representation and a representation of some condition **Y** iff: (a) that **X** correspond by a certain rule with **Y** is a necessary condition on the performance, in accordance with a Normal explanation, of the proper function of the consumers of **X**; (b) **Y** admits of variations, within the class of representeds into which it falls, which accord, one-to-one, with a set of rule-governed transformations of **X**, within the class of representations into which it falls; (c) the rule-governed transformations of **X** have as a Normal condition of their proper function the corresponding variation of **Y**.

A Normal explanation relative to a specific proper function is, in turn, defined as an explanation of the manner in which that function has historically been performed by the organ/system which possesses it (Millikan, 1986, p.52).[69] A condition which is necessary for the performance of a certain function in accordance with a Normal explanation is a Normal condition (Millikan, 1989, p.287). Hence representational status and content is determined in cases in which a rule-governed mapping between representation and represented is a Normal condition for the proper functioning of the representation consumer (Millikan, 1989, p.287).

Conditions (b) and (c) specify what is sometimes referred to as Millikan's isomorphism requirement (Neander, 2004; Shea, 2010), since the one-to-one correspondence between the variations in the represented and the transformations of the representation sets up a mathematical isomorphism between the class of representeds into which **Y** falls and the class of representations into which **X** falls.[70] Since this requirement will form a central point of

[69] As I understand Millikan, Normal explanations, in order to *explain*, describe certain features *under specific descriptions*, such that these descriptions pick out those properties of objects/events/circumstances, the instancing of which was historically causally relevant to the successful performance of a proper function. It is the features picked out by the explanations, rather than the features of the explanations which are relevant within her account. Hence, appeal to such explanations introduces a potential source of intensionality (with an 's') without rendering the account circular.

[70] It may, in fact, be better to construe Millikan as imposing a more general *homomorphism* requirement (see Shea (2010) and also n.96, below). Mathematically, a homomorphism differs from an isomorphism only in that the former need not be a bijection. Hence, all isomorphisms are homomorphisms, but not *vice versa*. Imposing this more general requirement would allow for two representations to have the same extension (cf. Shea, 2010);

focus in discussing **PNC**, **PIR** and **PER**, it would be worthwhile to explicate exactly what is meant here. A mathematical isomorphism is an injective and surjective mapping (i.e. a bijection), f:**G**→**H**,[71] which preserves the mathematical structure of the sets mapped. Hence, if we have an operation defined on **G**, say '*', and we have an operation defined on **H**, say '**.**', then for any g_1, g_2ε**G**, for any h_1, h_2ε**H**, $f(g_1*g_2) = f(g_1).f(g_2)$. That is, the operations performed on the elements of **G** correspond exactly to the operations performed on the mappings of such elements into **H**. In the same way, Millikan requires in condition (b), above, that the transformations which the representation can undergo correspond exactly to the transformations which its represented can undergo (cf. Shea, 2010), each within the class of representations and representeds to which they belong, respectively (Millikan, 1984, 2004).

My understanding of Millikan is that she holds that there are certain objective constraints on the variety of isomorphisms which can obtain between classes of representations and classes of representeds. However, out of the possible isomorphic mappings, that mapping which defines the representational content is determined by the needs of the consumer of the representation (Shea, 2010). Both the set of variations of the represented mapped, and which correspondence rule maps the representation onto that represented, consequently fall under the concept 'mapping in accordance with a specific rule'.

Hence, the system which *produces* the representation (the representation producer) has as its proper function to relate the representation with a certain item in the world and establish an exact correspondence between certain transformations of the representation and certain

---

this weakening of Millikan is, therefore, implicitly assumed in the discussion of **PNC**, below. Nevertheless, for ease of discussion, in what follows I will bracket this qualification.

It should also be noted that a further weakening of Millikan may be required: In order to accommodate the case of ambiguous representations, we may need to allow that the correspondence rule between representation and represented is not a *function* but is simply a *relation*. Again, this will be implicitly assumed in section VI when I discuss how **(TRP_{INF})** can accommodate ambiguous representations. However, for ease of discussion this qualification will be bracketed as well.

[71] That is for every distinct element yεf(**G**) there is only one distinct element in xε**G**, such that $f(x)=y$, and **H**=f(**G**).

variations of this item, under the direction of the requirements of the proper functioning of the representation consumer in accordance with a Normal explanation (cf. Millikan, 1984, 2000). Proper functions of the representation consumer and representation producer are, therefore, intertwined.

4.2.1. Proposed Solutions to the Problems of Determinacy

The paradigmatic examples in Dretske (1986, 1988) fall out as a very special case of **(TRP)**, since they are special in three senses. First, the operations on the set of representations and the set of representeds are each the respective identity operations on these sets. Second, the correspondence rule between representation and represented is of a distinctive sort (indication). Third, the same state which adapts the consumer to the environment is responsible for initiating behaviour in accordance with that adaptation.[72]

Hence, indication is *one* way in which a type of internal state can correspond to an external state type, but it needn't define the correspondence rule between representations and representeds which partially constitutes that mapping which is a Normal condition relative to the consumer's function. 'Normal' here does not correspond to *statistically* normal, and representation and represented may need to co-occur only infrequently to contribute to the fitness of the organism for precisely those reasons highlighted in the discussion of producer-based teleological accounts, above (Millikan, 1989).

HP

In addressing **HP** a proponent of (**TRP)** must rely on the idea that on this account the content of a state is not determined by its *causes* but by the relation between its correspondence to the world by a rule and the proper functioning of its consumers (Millikan, 1984). Hence, a representational state will represent one rather than another of the items appearing on the

[72] See Millikan (1984) for the definition of 'adapt'.

horizontal axis just in case that the representation maps onto *this* item (rather than any other) by a certain rule is a Normal condition for the consumer's proper function.

It would seem, then, that **(TRP)** can handle **HP** in a more satisfactory way than the previous accounts which we have considered. However, the actuality of this potential success is contingent upon: (i) whether there is a determinate item on the horizontal axis onto which the representation must map in order for the representation producer to perform its proper function in accordance with a Normal explanation;[73] (ii) whether the horizontal contents generated by this account are intuitively the correct ones.

With regard to (i), **(TRP)** must adequately address the possibility that there may be more than one representation consumer and, hence, *prima facie*, more than one horizontal item the mapping onto which constitutes a Normal mapping.[74] A difficulty more specific to **HP** is that **(TRP)** may entail that there is *no* item on the horizontal axis which is represented by the representational state. The horizontal axis is constructed out of the more proximate and more distal *causes* of the occurrence of **X**. However, **(TRP)** does not require that the property which is Normally mapped by **X** must be causally relevant to the production of **X**s.

This feature of **(TRP)** can be seen as a virtue, since it allows for the attribution of content (like *food*) which applies to properties which, *prima facie*, are not themselves causally relevant to the production of a representational state. However, it may also be a vice. Ultimately, *whether* this is a problem, and if so just *how much* of a problem it is, for **(TRP)** will depend upon whether **(TRP)** also generates highly unintuitive contents, and what the prospects are for constructing an alternative account which is able to yield contents like *food*. This issue will be explored further in the discussion of Millikan's response to the problem of complex causal roles and will be taken up again in the final section.

[73] Hereafter, that mapping between a representation and an item which is a Normal condition for the proper functioning of the representation consumer will be referred to as the Normal mapping.

[74] This issue will be discussed in more detail below.

VP

In addressing **VP** the proponent of **(TRP)** must rely on the idea that the representational state will have content ***P*** rather than ***Q*** or ***R*** (where <***R***> is a subset of <***P***> which is a subset of <***Q***>) just in case that it maps onto the property expressed by ***P*** by a certain rule is a normal condition for the proper functioning of the representation's consumer. On the other hand for those problematic disjunctive properties which can arise for **TRP**, the solution in the offing is that since one of the disjuncts is not causally efficacious in the execution of the proper function of the representation consumer, the mapping onto this disjunct needn't appear in the Normal explanation relative to the consumer's function. Hence, it should not appear in the content of the representation.

The crucial difficulty for this proposal has already been mentioned in relation to **HP**: Given that there is usually more than one consumer of a representation, will content always be suitably determinate? On what basis are we to be assured that the normal explanations relative to each of the proper functions intersect, and intersect uniquely relative to the mapping conditions which they mention, so as to yield a shared Normal mapping condition, thus yielding sufficiently determinate mental content (see also Neander (1994))?

Two of the central contributions of Millikan to naturalistic semantics are the differentiation between representation producers and representation consumers and the positing of the relation between them as central to content determination: The class of representation producing/consuming pairs which are relevant to the determination of representational content is the class of *cooperative* systems, systems which have co-evolved because the proliferation of the one depends upon the proliferation of the other (Millikan, 1984, p.97; Millikan, 2004, pp.73-74). Hence, we might extend this reasoning to suggest that the success of the organism depends upon the *co-ordination* of the representation consumers as they execute their proper functions, *in response to the representation produced* (cf. Millikan, 1989). Hence, Millikan might argue that were the Normal mappings relative to each

consumer such as to require that the representation simultaneously mapped onto *different* features of the environment, then this would undermine cooperation between the various consumer systems and hence would tend not to be favoured by natural selection.[75]

However, there are two central problems with this argument. First, it assumes that the existence of *different* Normal mappings (where the difference in each case is, at least, a difference in the object mapped onto) relative to different representation consumers necessarily implies an *incompatibility* between the Normal performance of the consumer proper functions[76] which require these Normal mappings. However, this is a fallacy. Just so long as the relevant mappings can be simultaneously instantiated when the Normal proper functioning[77] of the organism requires the simultaneous Normal performance of the proper functions of more than one representation consumer, there will be no incompatibility of function. It is, moreover, plausible that such simultaneous instantiation is possible: The case of properties which are locally co-instantiated in the environment in which the organism evolved would provide such a possibility.

Second, even if we were to grant this questionable assumption, an incompatibility of proper functions performed simultaneously *in accordance with a Normal explanation*, does not imply that the proper functions cannot be simultaneously performed *simpliciter*. Millikan (1986, p.52) gives the example of certain metabolic processes in the human which, though they *can* be performed in the presence of certain lithium compounds, nevertheless have historically been performed in the presence of calcium. Hence, that not all of the proper functions of the representation consumers can be simultaneously performed in accordance with each one's Normal explanation does not imply that the fitness of the organism of which the consumers form part is necessarily compromised, nor that the functions cannot be performed, nor that they cannot continue to be selected for.[78] Hence, content can be

[75] I am not aware of any published source in which she makes this argument, however.

[76] That is the successful performance of these proper functions in accordance with a Normal explanation.

[77] See n.76.

[78] Notice that this case would not imply that the Normal explanation, and hence the content of the representation, would change. So long as the Normal explanations describe Normal mappings which are

indeterminate due to the existence of multiple distinct Normal mappings, without thereby prejudicing the fitness of the organism/species of which it forms part. Hence, there is no simple argument from the need for co-ordination in consumer functions to effect fitness, to the conclusion that content must be suitably determinate.

On the face of it Millikan's account doesn't seem to be able to help with either **PNC** or **PIR**. *Prima facie,* correspondence by any rule to water is necessarily correspondence by the same rule to H_2O. Similarly, in the case of necessarily co-instantiated properties, such as the Quine-type properties, it would seem that any naturalistically specified correspondence rule obtaining between a representation and one of these properties would be a correspondence rule obtaining between a representation and another of these properties. Moreover, even if there could be naturalistic correspondence rules which obtained between the representation and one, rather than another, of these properties, since natural selection cannot separate necessarily co-instantiated properties and, *de facto,* mappings of such properties, there would seem to be no differential selectional advantage to the consumer between these mappings. Hence, *prima facie* **(TRP)** cannot give a non *ad hoc* reason why a representation should be counted as a water representation rather than an H_2O representation. Nor can it give a *naturalistic* reason to differentiate representations of Quine-type properties.

However, this overlooks the value which conditions (b) and (c) bring to the account. Indeed it seems to me that the kind of response to **PNC** and **PIR** which is available to a proponent of **(TRP)** is the most promising of all of the possible lines of response which we have reviewed in the present paper. In order to access the response, however, we must delve a little deeper into the theoretical resources which Millikan (1984) develops.

Given the conditions placed on representation in **(TRP)** Millikan defines two important elements involved in meaning and representation: The Fregean sense (or just 'sense') of a

incompatible, but not every subset of such Normal explanations have this property, then content would be indeterminate, whilst each consumer could perform its proper function in accordance with a Normal explanation enough of the time to ensure stability in the Normal explanation over time.

representation and the referent of a representation. The former is the Normal mapping rule (Millikan, 1984). The latter is that onto which the representation Normally maps (Millikan, 1984). Millikan also introduces one final aspect of meaning which, although not part of the definition of a representation in the way in which the two terms just introduced are, nevertheless can be accommodated within her account: Explicit and implicit intensions. Both of these notions apply to that which governs the tokening of a representation (what Millikan (1984) calls its iteration). An explicit intension is the iteration of a representation, **X**, according to specific rules or functions which take as arguments prior representational structures which do not contain **X**.

To give an example (one of Millikan's (1984)) suppose a primitive creature has amongst his set of representations one (**RED**) expressing the property of redness, one (**ROUND**) expressing the property of roundness and one (**FISISE**) expressing the property of being fist-sized, but no representation expressing the property apple. Apples are in fact highly nutritious for the creatures in question, and over time, as the creature evolves, a representation, **APPLE**, develops such that this representation is tokened according to the rule that whenever the string **RED ROUND FISISE** is tokened then so is **APPLE** and the latter representation replaces the former string in all further representations. That **APPLE** maps onto apples according to a certain rule becomes a Normal condition for the performance of the proper function of the consumers of **APPLE** and hence **APPLE** comes to express *apple*. **APPLE**s have as their explicit intension **RED ROUND FISISE**.

Three points which Millikan makes are very important for our purposes. First, both a representation and its intension may have senses, yet not have the *same* sense. Second, the referent of an intension and the referent of the representation which has this intension can come apart as well. Both of these points are generated by the fact that different mappings may be required for (orthographically) distinct representations in order to facilitate the proper functioning of the consumers as they respond to those representations. From these first two points it is striking that intensions act very much like reference-fixing descriptions (a point that Millikan (1984) herself notes), in the sense that they introduce a term into a language without being semantically equivalent to that term. Third, two representations may have the

same referent but *different* senses. The Normal proper functioning of the representation consumer determines which of the many isomorphisms which obtain between a set of representations and a set of represenTeds determines representational content. Hence, two different representations can map Normally onto the same referent in different ways (Millikan, 1984), since the role which these representations play in their respective consumers' (which needn't be the same) Normal proper functioning may differ.

Let's return then to the two problems with which we are concerned. There are two distinct ways in which the proponent of **(TRP)** may respond to the problem of **PNC**. The first is to point out that it is possible for two representations to have the same referent but different senses. It is true that there can be no selectional advantage between corresponding, according to a certain rule, to water and corresponding, according to the same rule, to H_2O. However, there can be a differential selectional advantage between mapping onto the property that is water and H_2O *in one way* rather than another (i.e. in having one sense which maps the representation onto water/H_2O rather than another).

Could the Normal mappings of two (orthographically distinct) representations and *water* differ in the way required? To answer this question fully we would need to know how the notion 'transformation' can be generalised. If it applies only to cases in which syntactic structure is kept invariant whilst lexical content is varied (Millikan, 1984), then it would seem that two mappings could certainly differ in the way required. For example, if being able to manipulate the chemical structure of H_2O can yield selective advantage, then the representation consumers which are Normally responsible for this can determine such a mapping. However, this interpretation of 'transformation' renders the current response as equivalent to the Dretske/Fodor line discussed above. We found there that the difference identified by the complex/simple distinction wasn't *the* difference after which we sought.

Nevertheless, the fundamental notion of an isomorphism is suitably abstract enough to allow the possibility that 'transformation' is not limited to this interpretation. Hence, this line of

response is not necessarily subject to the same objections which we have raised for the other accounts.

Despite this, the objections to the Dretske/Fodor line suggest strongly that the fundamental, broadly semantic, fact which underpinned the felt difference between water representations and H_2O representations was intimately connected to features of how the representation was acquired. Indeed, in the case of Sam, whilst we intuitively wish to restrict the *content* of Sam's relevant states to the single property expressed by those states, the felt difference between them derives from how these representations were introduced into Sam's inner 'language'. Millikan's notion of an intension can capture this nicely.

It also captures two of the central features of the standard Kripke-Putnam story which generated problems for the previous accounts in relation to the semantics of natural kind terms. First, it captures the notion of reference fixing descriptions, and Millikan's notion of an implicit intension[79] promises to capture reference fixation which is not dependent upon other representations. Second, it captures the distinction between reference fixation and semantic equivalence. Notice that **(TRP)** has a distinct advantage over informational-based semantics here. A representation, **X**, introduced into the subject's 'inner language' by a combination of other representations will potentially represent some item on an informational based semantics only if this combination of other representations carries information about that item. However, the difficult problem then arises as to how it is possible for **X** to represent this item whereas the combination of other representations does not, i.e. how is it possible for **X** to be semantically distinct from the combination of other representations?

(TRP) is able to cleanly deal with this problem because on this account what determines content is the proper function of the consumers of the representation, and there may be important differences between the intension and **X** in this respect. Hence, on **(TRP)** the

[79] This notion refers to the iteration of representations according to rules or functions which take as arguments perceptual data (1984, p.148).

combination of representations which introduces a distinct representation into a system needn't be semantically equivalent to **X**.

Hence **(TRP)** is able to non-arbitrarily exclude the reference fixing intension from the semantic content of the representation the reference of which is fixed. Yet it is able to non-arbitrarily differentiate between water representations and H_2O representations: These two representation types differ because either (i) the perceptual abilities involved in their implicit intensions differ, or (ii) the descriptions used to fix their references differ. This accords with the counterarguments given to the alternative theories above, and with central features of the standard story about natural kind semantics.

PER

It will readily be noted that this approach promises to deal nicely with **PER** as well. It was noted above (in the discussion of informational semantics) that: (i) It is, in many ways, plausible that empty representations seem to have content, and, hence, to differ from one another, because they are constructed out of other representations which are not empty; (ii) that this *prima facie* plausibility is undermined by the fact that there are empty simple representations (e.g. phlogiston). The present approach seems to 'thread the needle', as it were, by allowing, first, that an empty representation can be *introduced* by a complex representation. Two empty representations may, therefore, differ in intension whilst being identical in content (that is, because they have no content). This promises to explain the felt difference between two empty representations in parallel with the second response to **PNC**, above. However, second, this approach allows that the empty representation, though *introduced* by a complex representation, is, nevertheless, not semantically *equivalent* to it. Hence empty representations may be both syntactically and semantically simple. Since the model by which it achieves this result seems to capture certain features of natural kind semantics so well, this would explain the phlogiston case very nicely.

Let's turn finally to **PIR**, which has, in many ways, proven to be the most intractable of all the problems with which we have dealt. It has been noted above that it would seem that there

can be no selectional advantage to having states which express, say, *rabbit* and states which express, say, *undetached-rabbit-part*. Equally, in parallel with the above discussion of Dretske (1981), there are no explicit intensions to which we can appeal in addressing this problem. Hence, the work in solving **PIR** falls on the shoulders of the *primary* machinery of **(TRP)**.

Notice first that it is not clear that it is actually true that there can be no selectional advantage in having states which express *rabbit* and states which express *undetached-rabbit-part*. Consider the example of the bear. Gende, Quinn and Wilson (2001) have documented the phenomenon of selective consumption of salmon parts: That is the phenomenon of the bear consuming only certain parts of the salmon such as the eggs (in the female) or the brain (in the male) and then abandoning the rest of the fish. Gende, Quinn and Wilson (2001) relate this sort of behaviour to the level of availability of salmon in the bears' habitat during feeding periods: Roughly, the greater the density of available salmon, the more selective the bear's consumption of it. There is a clear survival advantage in this pattern. When the amount of available salmon is high, the bear can eat to satiety on the energy rich female salmon eggs and male salmon brain. In so doing the bear ingests more energy for one stomach-load of food, than it would if it consumed only whole fish. When it comes time to hibernate for the winter, the bears which have been consistently filling their stomachs with fish eggs are likely to do better than those which have been eating only whole fish.

Just what the representational capacities of bears are is unclear (to me at least). However, if bear behaviour *could* be explained with reference to representational states, then it seems highly plausible that the bear which could represent parts of the fish would have a selective advantage over the bear which could not.

Now, this doesn't show quite what we were trying to, because the problematic contents in **PIR** are not *part of a rabbit* and *rabbit* (these are not necessarily co-instantiated), but rather *undetached part of a rabbit* and *rabbit* (these are necessarily co-instantiated). Hence, to deal with this complication, consider the following hypothetical scenario. Suppose that in some

world the salmon are such that when the part containing the eggs of the female, or the part which is the brain of the male, is attached to the salmon it is such that if consumed from that state it is just as nutritious and energy-filled for the bear as it is in the actual world. We suppose that this is true because bear saliva can maintain the nutritious state of the part of the fish. However, if these parts are first detached from the salmon (say by a tempestuous fish eagle) and *then* consumed by a bear, they are poisonous to it. Supposing that the same story applies as before about the advantages of consuming only certain parts of the salmon, it would seem that it is advantageous to the bear to consume only certain undetached salmon parts. We might also generalise on which parts are particularly nutritious to the bear by supposing that just which parts are particularly energy-filled change each year and the bear is capable of adapting to these changes. Then it would seem that, intuitively, the bear which could represent *undetached part of a salmon* would be at a distinct advantage to the bear that could represent *salmon*. Furthermore, allowing the density of fish to change over time, it seems that the bear which could represent both *salmon* and *undetached part of a salmon*, would be at an advantage over the bear which could only represent *salmon*.

So far, so good. We've got to the point where if all of the above is correct we can say that *if* the bear could represent *undetached salmon part* and *salmon* this would be advantageous. However, perhaps many things which are advantageous are not possible. Given the argument with which we began discussion of **(TRP)**'s approach to **PIR**, how *could* the bear represent salmon.

Notice first that the aforementioned argument can now be seen to underestimate the power of the mapping conditions in **(TRP)**. For the mapping condition does not (or does not *just*) concern a function which obtains between two elements (a representation and a represented). Rather it has three features: (i) an operation on a set of representations, concerning a subset of the transformations which they can undergo; (ii) an operation on a set of representeds, concerning a subset of the variations of which they admit; (iii) a function between these to subsets which is an isomorphism between these two sets under these respective operations. It is these *three* features which define the mapping with which **(TRP)** is concerned.

Hence, if that the pattern of transformations which a representation can undergo corresponds in a certain way to the pattern of variations which one (out of a number) of possible representeds can undergo is such that the exploiting of *this* correspondence historically facilitated the production of an effect by the representation consumers which was selected for, then it is *that* correspondence rather than some other which determines what the representation represents. Hence, it is true that any function (in the sense of element (iii)) between a representation, **X**, and *salmon* is a function between that representation and *undetached salmon part*. However, some of these functions will preserve a certain correspondence between certain transformations which **X** can undergo and certain variations which *salmon* (or its instantiations *qua* such instantiations) can undergo. Other functions will preserve a certain correspondence between certain transformations which **X** can undergo and certain variations which *undetached salmon part* (or its instantiations *qua* such instantiations) can undergo.

Thus, we can frame two sufficient conditions for a representation to express *undetached salmon part* rather than *salmon*:

(**RS)** A representation **X** expresses *undetached salmon part* rather than *salmon* if:

(i) there exists a function which preserves a correspondence between certain transformations on **X** and a set of variations which instantiations of *undetached salmon parts* (*qua* such instantiations) can undergo, but which instantiations of *salmon* (*qua* such instantiations) cannot;

(ii) this correspondence is such that the successful performance of the proper functions of the representation consumers as they responded to the representation historically depended on it.

It will be readily noted that we can generalise over ‘salmon’ and ‘undetached salmon part’ to the set of all Quine-type properties.

Let’s recap. Fodor’s (1994) central insight was that the metaphysical properties which an item possesses, in particular the relations which that item, insofar as it instantiates the

property, can bear to others, is central to separating out fine-grained contents in **PIR**. In particular, the mirroring of these relations in the inferential relations which a representation can enter into was claimed by him to be central to this task. However, Fodor's suggestion fell short because, first, it could give no account of why certain metaphysically necessary truths about the relations which the relevant properties could and could not enter into should constrain the set of inferences which a possessor of the representation could and could not make. Second, it could give no account of why certain inferential relations which the representation could and could not enter into should determine that the representation referred[80] to one thing in the world instead of another, rather than simply determining that one thing was thought about in two different ways. Third, it fell short because it could place no non *ad hoc* bounds on the network of inferences which were relevant to the determination of this content.

The present suggestion doesn't suffer from any of these problems.[81] According to **(TRP)** *nothing* represents unless it maps according to some rule, and it is this mapping which determines what is represented. Quine-type cases are not special in this regard. Hence, for every case of representation, there are operations, which govern certain events which are entirely internal to the cognitive functioning of the organism, which help to determine content. Which of the truths about the relations which the represented can enter into and which it can't are those truths which are relevant to determining content, is determined by which mapping between representation and represented is a Normal condition for the proper functioning of the consumer of the representation.

[80] 'Refer' is intended as a broad term encompassing both application to individuals and application to objects in a term's extension.

[81] It should be noted that Millikan (1984) thinks that inferential processes are parasitic on, rather than determinative of, content. Hence, she would not approve of inferential processes being construed as potentially defining the operation on the set of representations under which this set is isomorphic to the set of representeds. However, it doesn't seem to me that we need to go along with Millikan on this point so long as 'inference' is explicated in broadly causal terms and the representations are individuated orthographically. Nevertheless, it should be noted that my proposal of how **(TRP)** can deal with **PIR** is not *committed* to construing the relevant operation on the set of representations as inferential: *Any* operation which can mirror, under an appropriate correspondence rule, the relevant operation on the set of representeds will do.

Cases in which the same thing is thought about in two different ways are distinguished from cases in which two things are thought about as follows. **Y** is thought about in two different ways when there are two representations, **X** and **X***, which map according to *distinct* mapping rules, *M* and *M**, such that both of the following conditions are met:

(i) the operations on the class of representeds defined by the respective mappings are each satisfiable by **Y**;

(ii) that **X** map onto **Y** according to *M* is a Normal condition for the proper functioning of the consumers of **X**, {*C*}, and that **X*** map onto **Y** according to *M** is a Normal condition for the proper functioning of the consumer of **X***, {*C**}.[82]

On the other hand, there are representations of two distinct items **Y** and **Z** if the following condition is met: There are representations **X** and **X***, there are mappings *M* and *M**, such that, that **X** map onto **Y** according to *M* is a Normal condition for the proper functioning of the consumers of **X** (whereas that it map onto **Z** is not), and that **X*** map onto **Z** according to *M** is a Normal condition for the proper functioning of the consumers of **X*** (whereas that it map onto **Y** is not).[83] The above discussion shows how it is possible within a naturalistic framework for two mappings to differentiate between two Quine-type properties and how there could be a selectional advantage in such differentiation.

Finally, **TRP** doesn't run into the holism problems which Fodor's account does, because there is an independent constraint which determines which set of transformations of the representation are relevant to content, namely the Normal proper function of the representation consumer.

This, then, offers an, admittedly potential, but, to my mind, plausible theoretical solution to **PIR**.

[82] We allow here that {*C*} may be equivalent to {*C**}.

[83] Here it may or may not be the case that **Y** and **Z** can each satisfy both *M* and *M**.

V

The Problem of Complex Causal Roles

5. *The Problem and Proposed Solutions*

5.1. The Problem

Teleological approaches to mental content, therefore, attempt to resolve the determinacy problems of broadly causal accounts by appealing in some way to the proper function of traits (either of the representation producer or the representation consumer). However, Neander (1995) and Griffiths and Goode (1995) urge that the functioning of a given structure or trait need not (indeed cannot) be determinate in the requisite way. This is because an effect's causal contribution to the selection of the trait which is responsible for it is realized by a number of intermediary mechanisms (see Griffiths & Goode, 1995). In the classic case of the frog, for example, predatory mechanisms (tongue reflexes) are engaged by any projections of small, dark moving things onto the frog's retina (Neander, 1995; Sterelny, 1990). Since most small, dark, moving things in the frog's natural habitat are edible flies, the production of the internal state in a suitable relation to the presence of certain retinal projections: (i) contributes to the selection of the trait responsible for it, by (ii) contributing to the fitness of the species, by (iii) occurring in response to food, by (iv) occurring in response to flies, by (v) occurring in response to small, dark, moving things, by (vi) occurring in response to certain retinal stimulations (Griffiths & Goode, 1995; Neander, 1995).[84] This concertina (Neander, 1995) of functions, therefore, renders it indeterminate as to what the state represents on teleological accounts. This is the sixth determinacy problem with which we will be concerned (hereafter, **PCR**).

[84] See figure 8, below, for a diagrammatic illustration of **PCR** in the case of the frog. The diagram is based on that given in Griffiths and Goode (1995), with minor amendments.

5.2. Proposed Solutions to the Problem

A number of different responses to **PCR** have been given in the literature; so much so that reasons have been given for taking the content of the frog's perceptual state to be fixed as *small, dark, moving thing* (Neander, 1995), *fly* (Sterelny, 1990), *food* (Millikan 1991; Papineau, 1997) and *small, dark, fast-moving food* (Agar, 1993; see Griffiths & Goode, 1995). I will consider four types of approach that have been taken to the problem. Although none of them are ultimately correct each highlights an important aspect for a more satisfactory solution.

5.2.1. Appeals to the selection for/selection of distinction

Griffiths and Goode (1995) identify (and reject) one argumentative strategy (which they refer to as the 'target of selection argument') for discriminating the content of the frog's perceptual state. These authors suggest that each level of functional description in the teleological concertina of effects can be seen as corresponding to distinct levels of theoretical description. The argumentative strategy attempts to apply Sober's selection of/selection for distinction across theoretical levels so as to privilege one effect/function in the concertina over others. Let $\{p_i,\ldots,p_n\}$ denote the n levels of a concertina of functions of the kind described above. The argument proceeds by asking us, for some p_i ($i\varepsilon(1,n]$), to consider the counterfactual scenarios in which: (a) the p_j (j below i) is/are held constant but p_i is absent; (b) p_j is absent but p_i is held constant. The result of this counterfactual version of Mill's methods is that in (b) but not (a) the mechanism producing the perceptual state is selected; from this it is concluded that the proper function of the mechanism is that described at the level p_i.

This argument is flawed for two reasons. The first is that the argument overlooks the possible interaction between two causally relevant factors. One way in which causally relevant factors, **c** and **d**, can interact is in the case where one factor, **c**, can only make its contribution *given that* the other, **d**, makes its contribution. This sort of case is sometimes referred to in the literature on experimental design and analysis as a catalytic interaction between two variables/factors, and, indeed the case of a chemical catalyst provides a nice illustration of it.

Suppose that two chemicals, *p* and *q*, react to form the chemical *r*. It may be possible that there exists a fourth chemical *s* such that the addition of *s* to the mixture of *p* and *q* can increase the rate of reaction from that which would occur if *s* were not added to the chemical mix. However, *s* may be such that *on its own* (i.e. in the absence of *p*) it will not react with *q* to form *r*. Suppose that not knowing any of these details, on an occasion in which *p*, *q* and *s* are all mixed together, we wonder whether it is *p* or *s* which is causally responsible for the production of *r* that is observed. Dutifully applying a counterfactual version of Mill's methods as used in the target of selection argument would yield the result that it was *p rather than s* which is causally responsible for the production of *r*. However, clearly this would be incorrect: *s* does make a causal contribution to the particular production of *r* which is observed; however, it can only do so in the presence of *p*.

In any such case, therefore, considering counterfactuals of the above form will lead us incorrectly to conclude that **c** makes no causal contribution. In the specific sort of evolutionary types of cases with which we are presently concerned, the contribution to selection of each lower level mechanism/function *depends on* the successful performance of the function at the level immediately above it. Hence, when we consider any successive pair of functions in isolation, the target of selection argument can be used to privilege *any* of the higher level functions over those at lower levels (see Griffiths & Goode, 1995).

The second flaw in the target of selection argument is that, for similar reasons, if the above argument shows anything then it shows that it is *possible* for a lower level function/effect not to make a causal contribution to the selection of the trait responsible for it, rather than that such a function/effect *did* not actually make such a contribution. This is so since the counterfactuals which the target of selection argument employs require us to consider what would have happened were the function described at the i-th level of the concertina still to have been realised, without the function described at the j-th level having not been. In such a scenario, presumably one would have to imagine an *alternate* function described at the level j in the place of that function which *actually* occupies this position. If the trait responsible for the i-th level function in the counterfactual scenario imagined were to be selected, then it follows that the *actual* j-th level phenotypic expression of this trait *needn't* have been

selected. The higher level function *could* have still been selected without it. However, it does not follow that *given* that the actual j-th level function/phenotypic expression *was* selected, the fact that the trait responsible for it *was* responsible for it made no causal contribution to the selection of this trait. There may be *many* intermediary functions described at the j-th level which could realize the function described at level i, compatible with all of these functions having causally contributed to the selection of the trait responsible for them in the world in which they appear on the j-th level.[85]

Ultimately, therefore, even when we consider matters in terms of selection *for*, all of the effects on the concertina are selected for.

Griffiths and Goode (1995) attribute versions of this argument to both Sterelny (1990) and Agar (1993). I think that Agar is guilty of a version of it; I think Sterelny uses it but irrelevantly; my focus will, therefore, be on Agar.

5.2.1.1. Agar

Agar (1993) proposes that in determining those environmental properties which are causally efficacious in the selection of a representational state[86], **S**, we need to consider not only those properties which are relevant to the *existence* of the state, but also those environmental properties which are relevant to the state having the *particular structure* which it does have: "…we should work out which properties we cannot change without modifying the exact structure of [**S**]" (p.10).

The fly being small, dark and fast moving is causally relevant to the selectional history of the *structure* of **S**, since if the fly were not such, then **S** would not have certain of the causal

[85] This is related to the stressed importance of considering historical contingencies in evolutionary explanations; see Fodor (2008), Sober (2010), Griffiths & Goode (1995).

[86] Notice here that Agar shifts the emphasis from the question of what property of an item is selected for to the question of which environmental feature does the selecting (is the force of selection (Agar, 1993)) of the trait.

propensities which it does have (its relation to external stimuli on the one hand and to perceptual faculties on the other), and it is partly in virtue of these propensities that it makes its contribution to the fitness of the organism. The property of being food is causally relevant to the *existence* of **S** if not to its structure. However, *fly* does not qualify for inclusion in the representational content, since slight, but significant, alterations in the essential property of the fly, its genetic makeup, if everything else was kept constant, would make no difference to the existence or structure of **S**.

Agar's proposal is valuable in three respects. (i) It highlights the idea that the proper way in which to account teleologically for one of the proper functions of the **S**-producing mechanism being to produce **S**s in response to more proximal stimuli is in terms of the selectional history of (certain aspects) of the *structure* of **S**; (ii) It highlights the idea that, owing to the frog's cognitive simplicity, relatively undifferentiated contents are often appropriately attributed; (iii) it grounds what I take to be the right sort of account of how the semantic properties of a contentful state can be relevant to causal explanation of the state's effects (see Nicholson, 2009); (iv) it gives a useful way of identifying the aspect of the environmental which should feature in the content of the state, namely searching for those features which exerted selectional pressure on the state's structure and existence.

Nevertheless, Agar errs in the way in which he attempts to apply his proposal, for much the same reasons as those given in section 5.2.1. In this particular case, Agar errs in overlooking the phenomenon of co-evolution between frogs and flies (Griffiths & Goode, 1995), and hence in eliminating *fly* from the content of **S**. Agar also errs in delimiting which section of the concertina is content relevant. Since food will not contribute to selection under certain conditions (e.g. if the food is poisoned), a description of these conditions should feature in the content of **S** (Papineau, 1997). On the other hand, since **S** will not occur unless small, dark, moving things cast suitable projections on the retina, a description of these projections should be included in the content of **S**.

On reflection, one can see that the application of the selection for/selection of distinction to the simultaneous determination of the environmental features which are selectionally relevant to the structure of a state and those features which are selectionally relevant to the existence of a state cannot be satisfactory in principle. For, it will always yield a conjunction of the most proximal and the most distal environmental features which are mentioned in the descriptions on the concertina of proper functions, whilst excluding those features which are mentioned in the descriptions on the levels in-between. Given that the structure of a representational state, **S**, as intimately connected to the relation of **S** to sensory stimuli on the one hand and perceptual/cognitive/behavioural systems on the other, changes in the proximal stimuli associated with an object with which interaction is biologically beneficial to the subject of **S**, will alter the structure of **S**. On the other hand since it is the biological benefit itself which is relevant to the existence of **S** (with or without the exact structure which it actually has), changes to the contribution which **S** makes to the selection of the organism will alter the fact of **S**'s existence. Since, this contribution is reflected in the top-most level of description, however, it seems that as long as we hold these two features constant, on Agar's proposal, we can change the environmental features which are described at any intermediate level of the concertina as much as we like without affecting either the structure or the existence of **S**. Hence, the fundamental difficulty with Agar's proposal is that it does not show how representation of the items which feature at these intermediate levels of description is possible.

5.3. Appeals to the notion of malfunction

Neander (1995) holds that:

(i) All of the effects of some trait, **T**, in a concertina of the above form constitute functions of **T**;

(ii) However, we can still identify *the* proper function of **T** as that function which is specific to it.

(iii) It is in terms of this specific function that representation and misrepresentation can be explicated.

Neander (1995, p.114-120) begins with Robert Cummins' notion of a functional analysis, in which a complex system is decomposed into smaller functional units, the operation and the contribution to the whole of which can then be successfully described. This process continues recursively on each subsystem. For each smaller subsystem, that subsystem will have a correct functional description at the level of analysis where it appears explicitly and as an unanalysed component. It will also have a correct functional description at the level where it appears implicitly as a component of the larger system. However, each subsystem will only be considered to be *malfunctioning* when it fails to produce its biologically normal contribution to the functioning of its immediate supersystem. Hence, a subsystem will be considered to be malfunctioning when it fails to produce its biologically normal effect at the lowest level of description/analysis at which it appears explicitly and as an unanalysed component (Neander, 1995, p.129). Let's call this proposal about malfunctioning **NPF**.

In the case of the frog, Neander (1995) takes state **S** to be "whatever state the frog's optic fibres and associated neural processors are in when they cause tongue snapping" (p.130), and, by an application of the above principle, renders the proper function of the detection device of which **S** forms part as detecting small, dark, moving things. **S**, therefore, has the content '*small, dark moving thing*'. Hence, **S**s correctly represent when caused by a small, dark, moving thing; **S**s misrepresent when caused by something else. We can frame this proposal about misrepresentation as follows:

> **(NPM)** For any system, **Q**, responsible for producing a representational state, **S**, **Q**'s tokening of **S** is a misrepresentation if and only if **Q** malfunctions (according to **NPF**) in producing **S**.[87]

It seems to me that Neander has identified one genuine type of misrepresentation. However, there are two central problems with the proposal, both of which Neander notes (1995, p.130-136), neither of which she adequately resolves. First the account is too restrictive: On its own it always yields proximal contents. This is because, given a concertina of proper functions,

[87] That is, the production of **S** may partly constitute the malfunction of **Q**, or it may just be the result of the malfunction of **Q**.

$\{p_i\}_{i\varepsilon[1,n]}$, of the detection device of which **S** forms part, for every p_k, $k\varepsilon(1,n]$ there is a proper function, p_j, $j\varepsilon[1,n)$ such that the device does p_k *by* doing p_j. However, as the example of the frog illustrates, the successful performance of p_k depends upon more than the device successfully performing p_j; systems external to that of which the device forms part must also 'co-operate'. For example, successfully detecting flies depends upon more than successfully detecting small, dark moving things, since when a bee-bee pellet is flung across the visual field of the frog, the frog successfully detects a small, dark, moving thing which is not a fly (cf. Fodor, 1990a). Similar cases can be constructed for the substitutions of 'small, dark, moving thing' for 'fly' and 'retinal pattern of kind A' for 'small, dark, moving thing'. Hence, a failure of the device to successfully perform p_k does not necessarily imply a failure to successfully perform p_j; hence, p_j is a more specific function of the device than is p_k. Iterating for $k\varepsilon(1,n]$, yields the result that p_1 is the most specific function of the device and, hence, the item mentioned in the description at this level is the correct content for **S**.

The question consequently arises, how is distality then possible? This question divides into two. The first is the question of how *any* distality is possible. For example, once we take **NPF** and **NPM** seriously, it seems that the correct content for state **S** in the case of the frog is *retinal pattern of kind A*, rather than *small, dark moving*. The second is how the *right amount* of distality is possible. For example, in the case of the frog how is it possible to get contents like *fly* or *food*. The first of these questions will be addressed here; the second will be implicitly addressed in the discussion of the possibility of misrepresentation without malfunction, below.

In response to the first question, Neander (1995, p.135-136) claims that whilst her proposal doesn't give a principled answer to the problem of distality it doesn't preclude one either (p. 136) and that whilst her proposal seems to imply that the (proximal) items appearing at the lowest level of a concertina are always the contents of representational states it does not compel us to give this verdict (p.136). Her motivation for these claims is the idea that, in addressing the problem of distality, we can appeal to further principles compatible with that proposed, which would delimit a class of more distal features on which her principle of malfunctioning could be applied.

However, it is a little difficult to see how any further principle could be compatible with that proposed and still delimit a subsection of the concertina of functions $\{p_i\}_{i\varepsilon[1,n]}$ which excludes p_1. For the original proposal was to take the proper function which is specific to a subsystem (in this case a representation generating device) to be that biologically normal function which corresponds to the low*est* level of description/analysis at which the subsystem appears explicitly and as an unanalysed component. However, since, by the argument three paragraphs back, for any concertina of proper functions, $\{p_i\}_{i\varepsilon[1,n]}$, this function is always p_1, clearly any further principles which demarcate a class of functions in the concertina which does not include p_1 will not be compatible with Neander's proposal as stated.

Hence, Neander's suggestion must be that the combination of her proposal with certain further principles would yield the result that a subsystem is malfunctioning if it fails to successfully perform the lowest of its biologically normal functions *within some class*, $\boldsymbol{E_D}$. Here, $\boldsymbol{E_D}$ is: (i) a subset of the concertina of functions generated by the description of the subsystem when it appears explicitly and as an unanalysed component; (ii) determined by the further principles to which Neander appeals. Call this proposal about malfunctioning **NPF***. Hence, the bare proposal about misrepresentation in her (1995) must be understood not as she presents it but rather as follows:

> **(NPM*)** For any system, **Q**, responsible for producing a representational state, **S**, **Q**'s tokening of **S** is a misrepresentation if and only if **Q** malfunctions (according to **NPF***) in producing **S**.[88]

[88] See figure 8 for a diagrammatic illustration of the proposal in the case of the frog. The nature of $\boldsymbol{E_D}$ and **NPM*** illustrated in the diagram are as Neander intends them to be. Whether this intention can be realized, however, will be discussed below.

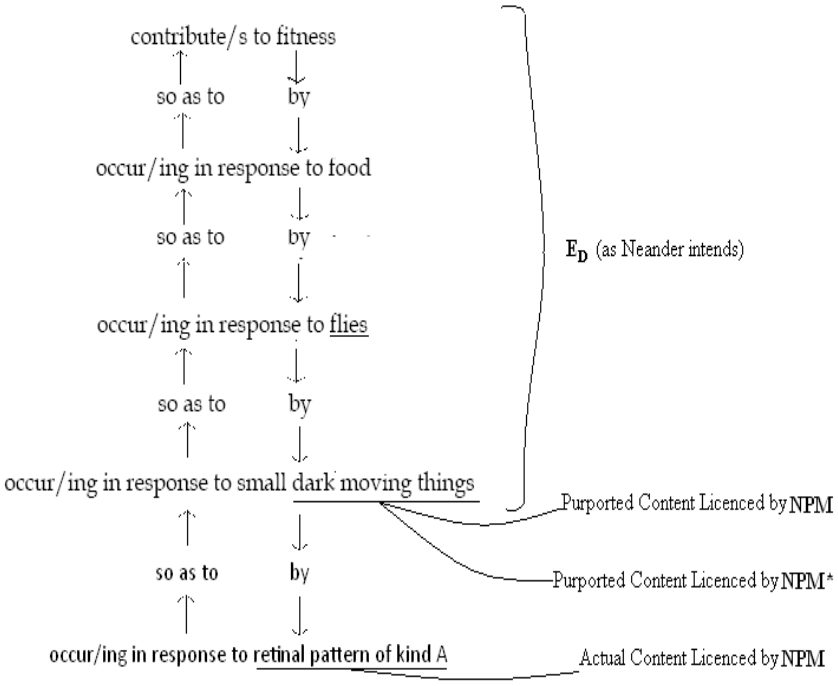

Figure 8

There are two problems with this move. The first is that Neander's proposal of prioritising the lowest level of functional description now becomes far less well motivated. I take Neander's central motivation for this prioritisation to be that it is intuitively plausible that a supersystem can malfunction without each one of its subsystems malfunctioning. This phenomenon occurs, moreover, in just those cases in which the possible causes of the non-performance of the supersystem's function are greater than one. In such cases all but one of the subsystems may be functioning in that way which normally[89] contributes to the supersystem successfully performing its proper function. Hence, the non-performance of a supersystem function will entail the malfunction of a subsystem if and only if it is *that* subsystem which is the causal source of the non-performance.[90] Neander's proposal, in turn, yields this result.

[89] Where this notion is not statistical but rather something like Millikan's notion of Normal.

[90] We can generalize this idea to the case in which two or more subsystems are causally responsible for the non-performance of the supersystem's proper function.

However, it is clear that this line of justification will only support **NPF** and, hence, **NPM**. For it supports the idea that there is a sense of malfunctioning such that a subsystem malfunctions only when the cause of the non-performance of a biologically normal supersystem function lies *entirely within* that subsystem. Clearly **NPM*** entails that it is possible for a representational subsystem to malfunction even if the causal source of the non-performance of the relevant biologically normal supersystem function lies outside of that subsystem. For it is only because of this entailment that **NPM*** has any chance of yielding distal contents.[91]

A similar theme permeates the three further subsidiary reasons which Neander gives for adopting **NPM**. First she suggests that a full account of representation and misrepresentation will need to depend upon: (i) the existence of inferential processes in perception; (ii) a two-factor theory which includes both sense and reference. **NPM** and its entailments about the content of mental representations make more plausible content attributions in the case of simpler organisms, allow for misrepresentation without malfunction via (i) in higher organisms, and inform (ii) by giving the "criteria by which a cognitive system identifies that which it represents, and how it represents it" (1995, p.134). However, it is apparent that these reasons for adopting **NPM** carry over to **NPM***only if the further principles appealed to in **NPM*** always delimit the class of proper functions such that the lowest level of description in the class is that level at which a description of the proximate environmental parameters

[91] Couldn't Neander claim that **NPM*** only entails that it is possible for a representational subsystem to malfunction even if the causal source of the non-performance of the relevant biologically normal supersystem function *extends* outside of (though still lies *partially within*) that subsystem? Alternatively, couldn't she claim that it doesn't even entail this, instead requiring that only those cases in which the representational subsystem fails to perform its lowest level function in $\boldsymbol{E_D}$ *because* it fails to produce its *absolute* lowest level function (i.e. that function described at level $\mathbf{p_1}$) are relevant? Not if she wishes to still maintain: (i) the equivalence between malfunctioning and misrepresenting; (ii) that her proposal addresses **PCR**. For on either of these two suggestions, it is a necessary condition for subsystem malfunction, and, hence, for misrepresentation, that the subsystem does not perform its absolute lowest function. Hence, if the subsystem *does* perform this function, then it *doesn't* malfunction and there is *no* misrepresentation. But how should the content of state **S** be construed to ensure this result? It could only be by construing this content as *disjunctive*, where the disjuncts are the items mentioned in the description of the $\mathbf{p_1}$-level function and the lowest function in $\boldsymbol{E_D}$, respectively, e.g. in the case of the frog, *retinal pattern A or small, dark, fast moving*. For, it is only in this case that misrepresentation requires that **S** be tokened in response to something that isn't mentioned in the description of the $\mathbf{p_1}$-level function. Hence, this proposal would be of no help in solving **PCR**, insofar as this problem arises within the field of naturalistic semantics. It would also be of no help in accounting for, what I take to be, the standard case of misrepresentation. In this case an item which is not in the extension of the content of the representational state causes this state's tokening *by* causing the occurrence of the *same proximal stimuli* that are normally caused by items which *are* in that extension.

occurs. This is a substantive assumption which Neander gives us no reason to believe and which, *prima facie*, needn't be true.

Finally, Neander suggests that **NPM** brings the account of mental content into line with Cummins' suggestion that "psychology should aim to provide an interpreted cognitive analysis" (Neander, 1995, p.135) which consists of a recursive decomposition of the cognitive system into parts according to the specific tasks which they perform. However, this reason can only support **NPM**, and not **NPM***, since it is only the former proposal which yields functions which are *specific* to subsystems.

In general, without a specification of the further principles which delimit the class $\boldsymbol{E_D}$, there is no guarantee that Neander's notion will do any substantive work in the final account of content and misrepresentation. Once we appeal to further principles to address the distality problem, there is no guarantee that the only satisfactory principles are such that they will not, *on their own*, yield determinate function ascriptions *without* appeal to Neander's sort of analysis. Hence, in order for her proposal to be accepted she would have to give us, not the further principles themselves, but rather what sort of work we have good reason to expect that these principles will do and what their limitations will be.

The second central problem with Neander's overall proposal is that a purported solution to **PCR** is bought at the expense of an adequate account of misrepresentation, since misrepresentation can occur without malfunction. In response to this Neander (1995, p.132) appeals to levels of cognitive complexity and the introduction of inference in perception formation. Neander's suggestion is that her account of misrepresentation applies directly only to the initial representations of environmental physical parameters; further instances of misrepresentation are due to errors in the process of inference on the basis of these initial representations.

It seems to me that this idea is a good one and introduces a further junction where error can occur. However, we can show that the success of this appeal *depends upon* a prior solution to **PCR** having been given, and, hence, cannot be appealed to in defending Neander's own solution to this problem.

In a naturalistic framework, the most plausible way to construe Neander's appeal to inference is as follows. The initial representations are taken as input values and a certain representational state is generated on their basis as the output value. Since the input values do not uniquely determine the output value, mistakes can occur.

Suppose, then, that the appeal to inferential processes in perception is intended as a way of accounting for the *fixation* of distal content. Arguably, on a naturalistic account of content inferential relations are most plausibly explicated in causal/functional terms, amended, perhaps, with an appeal to teleology in order to separate the constitutive functional relations from those that are not. Hence, state **S** occurs in response to a set of representations, $\{\mathbf{R_i}\}_{i\varepsilon[1,n]}$ of proximal stimuli, and the subsystem which produces **S** in such circumstances has the implementation of the operation describing this process as its proper function.

We want to say that **S** represents some distal item, **O**, because occurring in response to **O**s discharges part of the proper function of the subsystem of which **S** forms part. We also want to say that **S** misrepresents when it is produced in response to $\{\mathbf{R_i}\}_{i\varepsilon[1,n]}$ but in the absence of **O**. However, this does not constitute a malfunction of the representational system, since under certain conditions some non-**O**s produce $\{\mathbf{R_i}\}_{i\varepsilon[1,n]}$. But, it will readily be noted that the problem of **PCR** has just been introduced at one level higher: Since, **S** occurs in response to **O**s *by* occurring in response to $\{\mathbf{R_i}\}_{i\varepsilon[1,n]}$, in virtue of what does **S** represent **O**s rather than a disjunction of the elements of $\{\mathbf{R_i}\}_{i\varepsilon[1,n]}$? If we once again appeal to Neander's principle, then the content of **S** will be a disjunction of the $\mathbf{R_i}$ rather than **O**. If there is some alternative principle to which we can appeal, then what reason do we have to believe that it will not displace **NPM**/**NPM***?

Hence, Neander's appeal to inferential processes in perception must assume the prior existence of states, within a cognitive system capable of such inference, which have appropriate distal contents. Under these conditions, I agree that if human perceptual processes are inferential, then this would explain how misrepresentation without malfunction is possible. However, what account can be given for the fixation of this prior content? If it is a teleological account, then it too will face, and must solve, **PCR**. On the other hand, if there is some non-teleological account of content on which a teleological account amended with **NPM**/**NPM*** must *depend* in order to account for misrepresentation without malfunctioning, then it would seem that we would have adequate reason to abandon the teleological account so amended in pursuit of this alternative theory of mental content. For, the fact that the amended teleological account leaves explanatory gaps which an alternative account of content must fill, suggests that there are not really two adequate accounts of content here, but rather that the teleological account isn't really accounting for mental content.

5.4. Appeals to Representation Consumers

Millikan (1991, pp.162-163) argues that the content of the frog's representational state is *food*, since it is only when the state mapped suitably onto *food* that the way it has historically been used by its consumers allowed for the performance of their proper functions (Papineau, 1997). Hence, a mapping between the state and food would appear in the Normal explanation of the proper function of the consumer in responding to the state.

There are two problems with this solution. First, as Neander (1995) and Papineau (1997) both note Millikan's argument compels us to go further up the concertina: Just as historically it was only when the representation mapped onto food that the tongue reflex and digestive systems contributed to fitness, just so, it was only when the ingested food enhanced the health of the frog that the circulatory and cellular molecular processing mechanisms correctly performed their proper functions and contributed to frog fitness; and the argument recurs. We can see that this result readily generalises. Hence, Millikan's solution either: (i) cannot explain how it is possible for there to be representation of any item described on any level other than the top-most level of a concertina of proper functions, or (ii) must render disjunctive contents (where the disjuncts in the case of the frog would be '*food*' and the

relevant items mentioned on the levels of description above food in the concertina). Since, very often the items mentioned at the top-most level are not the intuitively plausible contents of representational states, and since it is surely *possible* for items appearing at lower levels of description to be represented, (i) would be unpalatable. However, (ii) would not resolve **PCR** as it appears in the field of naturalistic semantics. Hence, on either (i) or (ii), Millikan's solution is inadequate.

Second, Millikan motivates her approach to teleosemantics by emphasising that it is the *effects* of the representation producers on consumers' functioning which are relevant to natural selection, and not the causes of the production of such states. But divorcing contents from their causes, and fixing them in relation to consumer proper functions often yields intuitively implausible contents. For example Pietroski (1992) has argued that dimming light at nightfall may in some species trigger an internal representational state, **S**, which causally contributes to falling asleep. This effect causally contributes to the fitness of the species, partly, by protecting the organism from predators. Since one of the representation consumers is the system which brings about the falling asleep of the animal, and since this system has contributed to the survival and proliferation of the species, when it has, partly in virtue of protecting the organism from predators, it seems plausible that the mapping between **S** and *predator* is a Normal condition for the performance of the Normal proper functions of the representation consumer. Hence, on Millikan's account it would seem that **S** represents *predator* rather than *nightfall*, or *darkness*, etc., *contra* intuition.

5.5. Appeals to More Sophisticated Cognitive Abilities

Papineau (1997) attempts to utilize elements of both Neander's solution and Millikan's solution within his (1993) teleosemantic account. Papineau (1997) takes the relevant states of selection to be whole propositional attitudes, and suspects that in the case of cognitively simple organisms, like the frog, there is no fact of the matter about which of the concertina levels provides the correct content. He holds that it is the effect of *desires* which are content determining. On his (1993) theory a "desire's real satisfaction condition [is] that effect which it is the desire's biological purpose to produce" (p. 58-59). Similarly "the real truth condition of a belief [is] that condition which it is the biological purpose of the belief to be co-present

with" (p.59). For Papineau, belief content derives from desire content. Thus, Papineau's (1997) solution to **PCR** is to argue that the content of desires is determined by their *specific* effects (in Neander's sense) and the content of beliefs is determined derivatively from this specific desire content. The specific effect of a desire, in turn, is that effect which is: (i) relatively independent of the particular beliefs with which the desire combines; (ii) independent of the functioning of other nonrepresentational systems. Hence, it seems that this two stage algorithm is intended to define, in every particular case, two sets of effects, the intersection of which yields a singleton, which is the specific effect of the desire.

This solution is attractive for (i) the emphasis it places on the selection for effects, (ii) one type of possibility of combining Neander's analysis with other principles; (iii) the emphasis on differential levels of cognitive complexity in fixing content. However, ultimately Papineau's proposal is unsatisfactory. First, it does not adequately demarcate the two sets of effects, the intersection of which is supposed to yield a single determinate content. Second, the particular kind of cognitive complexity which it requires for determinate content is ultimately untenable within an evolutionary perspective which appeals to natural selection.

It is important to note that although Papineau seems to intend his proposed solution as a general one, he gives no reason to think that *in general* the two sets identified by application of his algorithm will have a singleton as their intersection, rather than a set consisting of more than one element, or no elements at all. In fact, we can show that there is good reason to think that in general Papineau's algorithm will *not* yield a singleton as the intersection between the two sets, and in particular that it does not do so in the specific case described by him.

First, it seems to me that the first stage of the algorithm can, in principle, yield a determinate set of effects in a way that is acceptable within a naturalistic theory of content. I take the best formulation of his basic idea to be that under conditions of proper functioning of the belief and desire systems, there either *is* or *can be* a wide variation in the particular beliefs (specified non-intentionally) with which the desire can combine such that a certain class of effects (call it $\mathbf{E_P}$) remains invariant under these variations. Hence, the notion of

independence of desire effects from particular beliefs, is underpinned by the idea that under conditions of proper functioning of the belief and desire systems (or, perhaps, better, the belief and desire types themselves) there will be a counterfactual dependence between the desire (specified non-intentionally) and a particular class of effects, such that were that desire to obtain then that class of effects would obtain.

The crucial question for Papineau's account, however, is whether it is possible for *any* element in $\mathbf{E_P}$ to be such that if it were not to obtain then this would entail a malfunction of the desire which caused it. That is whether there is any member of $\mathbf{E_P}$ which is independent of the functioning of other non-representational systems. I think that it is clear that the answer to this question in the particular case which Papineau describes has to be 'no'.[92] If we eliminate the set of distal effects through Neander's analysis, then we should also eliminate the effect of one's stomach receiving food: This effect also depends on other systems, namely the system responsible for the peristaltic movements of the oesophagus; this dependence, moreover, implies that should food not get into one's stomach this doesn't necessarily mean that the desire is malfunctioning.

Given the above discussion of Neander's proposal, we can, of course generalise this result. For Papineau's account is an example of an attempt to combine **NPM** with the additional principle that the desire must be independent of any particular belief. Applying the arguments presented against Neander to the present case, we see that, in general, we can expect the occurrence of any distal effect of a desire to depend on the contribution of factors lying outside of the immediate desire producing system. Even when beliefs and desires are functioning properly and the environment is similar enough to the one in which the most recent episode of selection of these cognitive structures took place, distal effects of a desire (i.e. the elements in $\mathbf{E_P}$) may fail to occur.

[92] In this example, he applies his algorithm to a system with a belief desire psychology in the attempt to show that the subject's stomach receiving *food* is the specific effect of a desire for food.

This is, in fact a feature which is readily recognisable from our belief-desire folk psychology. Different beliefs combine with a desire to cause different behaviours so as to facilitate the desire causing its proper effects despite changing circumstances (cf. Papineau, 1997). Though beliefs may inform the agent about aspects of the world relevant to the satisfaction of its desires, there are features of the world beyond the agent's ken which may thwart this satisfaction. These features moreover may be internal or external to the agent, but will concern either larger systems of which the agent forms a part or smaller systems which form part of the agent. Applying this to a biological context implies that the failure of occurrence of a distal effect of a desire does not intuitively entail that either the beliefs/desires on which the agent acted or the environment is abnormal.

However, this is not a satisfactory result as far as Papineau's analysis is concerned, because it entails that there is no member of $\mathbf{E_P}$ which is such that the failing to obtain of that member entails the malfunctioning of the desire. Hence, it seems, we must conclude that the set generated by the intersection of the two sets identified by Papineau's algorithm will, in general, be empty. If **PCR** generated too many contents, then Papineau's solution to it generates too few.[93]

What Papineau needs, in fact, to make his proposal successful is not just the notion of conditions of proper functioning, but rather the notion of conditions of *optimal* functioning, such that in optimal conditions the non-occurrence of a certain member of $\mathbf{E_P}$ entails the malfunction of the desire. However, *by hypothesis* all of the effects in $\mathbf{E_P}$ are effects for the production of which the desire type was selected. Hence, there is no *one* member of $\mathbf{E_P}$ relative to which we can fix the optimal conditions sought. Rather, we can define conditions, $\mathbf{C_{e(i)}}$ for each member of $\mathbf{E_P}$, e(i), such that under $\mathbf{C_{e(i)}}$, the non-occurrence of e(i) entails the malfunction of the desire. Which of these conditions will qualify as optimal for the

[93] Why can't Papineau adapt his proposal so that it comes out as an instance of **NPM***? Because, then we would need principled grounds for taking the specific effect of the desire to be the *lowest* element in $\mathbf{E_P}$, and neither Papineau nor Neander gives us such grounds. Of course the lowest element in $\mathbf{E_P}$ will be *more specific* to the desire than any other element in this set. But that is no guarantee that it will be very specific to the desire at all. So why should this *relative* specificity be what is required for determinate content? We would need a good deal of cogent argument from Papineau here, and it's not clear that such argument is possible in this case.

functioning of the desire type will depend upon the content of the desire (cf. Fodor, 1987, pp.105-106). Hence, the algorithm which Papineau provides for solving **PCR** either yields no content for representational states, or yields such content only by violating the non-circularity requirement on a naturalistic theory of content.

The second reason for the unsatisfactory nature of Papineau's account is that on this account the proper items of selection are entire propositional attitudes (or the ability to form such). But there seem to be good reasons for thinking that many such existing attitudes do not have or cannot have had a selectional history. In response to this Papineau seems to appeal to some form of (perhaps limited) compositionality (Godfrey-Smith, 1996; Papineau, 1993). However, that a complex structure has a selectional explanation does not imply that each part of it has such an explanation to call its own: Certain features of a complex structure may be necessary by-products of the collection and ordering of other features which are selected for and/or which are essential to the selectional explanation of the whole (e.g. the human chin seems to be a necessary by-product of the selection of jaw structures) (Gould & Lewontin, 1979). We need a principled reason for thinking that even though the proper object of selection is an entire complex structure, a desire, the relevant individual parts of it to which Papineau wishes to appeal have appropriate proper functions, and, therefore, content (see also Godfrey-Smith, 1996).

This suggests to me that a better strategy is to apply our teleological account to simpler semantic parts and to compositional rules. This, of course, makes it more likely that cognitively more simple organisms share certain representational capacities with cognitively more complex organisms (see also Godfrey-Smith, 2002). As has been suggested in section I above, this strategy is also supported on independent grounds.

VI

RESULTS AND AN ALTERNATIVE PROPOSAL

The above discussion displays a strong and revealing pattern. The biggest points of success (or at least strong promise of such success) for **TRP** lies in the resolution of **PNC**, **PIR** and **PER**. Importantly, this success derives squarely from (i) the teleological nature of the account, (ii) its consumer-based nature, and (iii) the natural accommodation of an isomorphism requirement. Hence, we cannot derive the same benefits by removing the determinative role of the consumer. However, this focus on the consumer generates, in certain cases, highly unintuitive contents.

These contents are not generated by information-based accounts. Moreover, such accounts, particularly when amended with a teleological graft, look appealing (despite their problems) in solving **HP** and **VP**. However, indication-based teleological accounts are not viable and it seems to be a consequence of consumer based accounts that the particular causal history of the representation is irrelevant to its content. Hence it would seem that consumer based teleology is incompatible with information based semantics.

Moreover, even if it were possible to blend these two approaches, the particular blend would still be left with the problems of multiple consumers and **PCR**. With respect to this latter problem, though each attempted solution reviewed above has value, none adequately resolves the problem.

The purpose of this final section is to attempt to do two things. First, to show that an information-based semantics is not incompatible with a consumer-based teleological one. Second, to sketch how an account blended from these two perspectives can: (a) address the problems of **HP** and **VP** whilst mitigating the problem of multiple consumers, and (b) address **PCR** in a way which utilises the insights of, whilst avoiding the difficulties associated with, each of the attempted solutions reviewed above.

6.1. Information and Informational and Consumer-Based Semantics Revisited

We have seen in sections III and IV, above that accounts which are based on a notion of information requiring a suitably high conditional probability of the instancing of the property expressed by a representation on the tokening of that representation face deep and severe difficulties. Indeed, on quite independent grounds, I think that it can be shown that an account of information such as $\mathbf{DI_1}$, in order to satisfy the motivating conditions for its adoption, must be amended with a broadly causal condition, in part of the form embodied in **FI**. Moreover, it can be shown that there is no sustainable weaker notion of information of the form embodied in $\mathbf{DI_2}$ which is amenable to inclusion in a satisfactory theory of mental content.[94]

Nevertheless, the attraction of an information-based semantics is undeniable. Consider, then, the following formalised notion of information, drawing on the accounts of both Dretske and Fodor:

($\mathbf{IC_N}$) For all **X**, for all **Y**, **X**s carry the information that **Y** if and only if:

(i) **Y**s *qua* **Y**s cause **X**s *qua* **X**s;

(ii) **Y**s have actually caused **X**s;

(iii) $0<P(\mathbf{Y},\mathbf{X})$.

Let us say that **X**s are *apt* for carrying the information that **Y** if and only if (i) and (iii) hold. It is questionable whether there is any theoretical work left for clause (iii) to do in a theory of information *per se*. However, if our interest lies in the construction of an information-related notion tailored to a naturalistic semantics then I think that (iii) earns its keep.

[94] The abandonment of $\mathbf{DI_1}$ and $\mathbf{DI_2}$ is partly motivated by the consideration, defended in sections III and IV above, that they are untenable as far as a naturalistic semantics is concerned. However, for details of the 'independent grounds' mentioned in the text, and for discussion of how these grounds can be seen to motivate the account of information adopted here, see Nicholson (2012).

Consider, then the following account of representation:

(**TRP$_{INF}$**) An inner state **X** is a representation of some condition **Y** iff:

(a) **X**s are apt for carrying the information that **Y**;

(b) that **X** is related by a certain correspondence rule, *R*, to **Y** is a necessary condition on the performance of the Normal proper function of the consumers of **X**;

(c) **Y** admits of variations, within the class of representeds into which it falls, which accord, one-to-one, with a set of rule-governed transformations of the class of representations into which **X** falls;

(d) the rule-governed transformations of **X** have as a Normal condition of their proper function the corresponding variation of **Y**.

As before, to say that **X** Normally *maps*, according to mapping, *M*, onto **Y**, encapsulates (b), (c) and (d).

In **(TRP$_{INF}$)** the mapping relation, *M*, is partially constituted by a conditional probability requirement stipulating that P(**Y**,**X**) = p>0. However, the exact value of p is determined by the needs of the consumer; that is p will depend upon the degree of probabilistic dependence between **X** and **Y** which is (part of) a Normal condition on the proper functioning of the consumers of **X**.

We should take a moment to pause over this point. Taking a leaf out of Millikan's (2004) book, we can say that the reference class for **X** is the set of all **X**s individuated by their proximal genetic origin;[95] the reference class for **Y** is simply the set of all **Y**s.[96] The conditional probability must be underwritten by some condition which explains its

[95] That is **X**s form a higher order reproductively established family (Millikan, 1984).

[96] This is one reason why it might be better to impose a homomorphism requirement rather than an isomorphism requirement: This weakening would allow for the correspondence function between representation and represented to be non-surjective. Nevertheless it doesn't *necessitate* the adoption of the weaker requirement so long as we can restrict the range of the correspondence function in a suitable way.

persistence, in line with Dretske (1988) and Millikan (2004). This is unproblematic, because the causal condition will help us to avoid the unwanted consequences that this interpretation of the correlation between **X**s and **Y**s has (as noted above).

Hence, the representation *consumers* determine *p*, since it is relative to their needs that certain states are selected (in part) for correlating in accordance with this conditional probability. It is, at least partly, by this mechanism that the conditional probability persists (cf. Millikan, 2004). Nevertheless, the *class* of possible representational states on which the proper functioning of the representation consumers 'acts' so as to determine the value of *p*, is delimited by the restriction that **X**s *qua* **X**s must be caused by **Y**s *qua* **Y**s.[97] Hence, the above account makes representation harder to achieve. This is because, on this account, representational content is a 'two-way' relation between organism and world. Not every physical system will be able to support this relation.

We are now in a position to see why a consumer-based teleological semantics is not incompatible with an informational semantics (properly construed). As Millikan (2007) notes since proper functions are *effects* which are selected for, it cannot be a proper function of a representation producing system to be *caused* in a certain way (see also Godfrey-Smith (1992)). Nevertheless, the fact that representation tokens have certain effects on their consumers, which effects are biologically advantageous from an evolutionary point of view because the representation *relates* to some external condition (i.e. because the effects occur in a certain relation to this external condition), can cause the representation type to be selected (Millikan, 2004; cf. Millikan's (1984) notion of a relational proper function).

Hence, consumer-based teleological accounts concern the organism-world direction of dependence, and only concern the world-organism direction of dependence insofar as it affects this first dependence (this is made clear in **TRP**'s requirement that the representation map onto the world). Hence, they do not preclude a further requirement (in this case a causal

[97] Consistent with this they may also be caused by **Z**s *qua* **Z**s.

requirement) being imposed to characterise the world-organism direction of dependence, unless this precludes any role for the representation consumer in determining the character of the organism-world dependence. Clearly **(IC_N)** doesn't do this.

The addition of the extra condition to **(TRP)** does not undermine its virtues in addressing **PNC**, **PIR** and **PER**. Let's then look at the prospects of **(TRP_{INF})** being able to offer us a satisfactory resolution of the remaining difficulties. The following diagrams will facilitate the discussion which follows.

6.1.1. ***HP***

Informational semantic theories which attempt to deal with **HP** by identifying the item on the horizontal axis which bears the appropriate counterfactual supporting relation to the relevant representational state type face the problem of disjunctive proximal contents. We have noted above that **TRP** promises to avoid this problem. On the other hand, an informational requirement like that embodied in **TRP_{INF}**, clause (i), promises to avoid the problem that the contents generated by **TRP** may not fall on the horizontal axis at all.

It was noted above that this is a problem for **TRP** if this feature of the account allows unintuitive contents to be attributed. The 'predator' example of Pietroski (1992), cited above, shows that this is indeed the case (as does the Kimu/Snorf example in Pietroski (1992)). What has gone wrong in each of these examples is that the state type which is tokened does not bear an appropriate 'sensitivity' to the condition which **TRP** attributes to it as its content.

On the other hand, **TRP_{INF}** yields the intuitively correct contents in these cases. For example, even if there is a counterfactual supporting relation which obtains between *predator* and the property of causing *some* internal state of the organism, this is not the state type which is tokened in response to dimming light at nightfall. Rather, the counterfactual supporting relation which obtains there is between *nightfall* or *dimming light* and the internal state.

Hence, if the creature represents anything it represents these two features of the world. Similar comments apply to the Kimu/Snorf case in Pietroski (1992).

Now, we have already noted that Millikan thinks that it is a drawback of associating the contents of a state with items which either do or can bear a certain causal relation to the state that this association will not allow one to get past these causal properties. Hence, she thinks that such an account cannot render distal contents like *food*, whereas an account like **(TRP)** can. Hence, the thought is 'so the much better for **TRP**'. However, it seems to me that on reflection we can equally well respond 'so the much worse for accounts such as **TRP**'. It is, of course, a virtue of an account if it can explain how an organism can acquire concepts such as that of food. However, it is a vice of an account if it renders this sort of content indiscriminately to simple and complex cognitive systems alike. The point is that the concept of food seems to be a reasonably sophisticated one. Hence, it seems implausible to think that a very simple organism (which is how the frogs or tortoises have been portrayed in the literature) actually has the ability to think about food *qua* food (cf. Neander, 2004). However, there is no principled reason on **(TRP)** not to attribute this content to simple creatures; and the same goes for any other sophisticated concepts, in particular abstract, theoretical concepts which refer to properties which make a difference to an animal's biological success.

In contrast, it seems to me that an account such as **(TRP_{INF})** can do a much better job here. My suggestion is that the possession of these sorts of concepts is dependent on reasonably sophisticated *existing* representational and/or non-representational features of the organism. For present purposes, let's assume that *food* is a functional kind, so that food is anything which has the disposition to maintain life and growth on ingestion. As the above definition indicates, functional kinds, like *food* are defined by certain of their characteristic *effects*. The concern, I take it, is that it is not the properties in virtue of which something (e.g. a fly) satisfies this functional description which are causally relevant in producing a representational state in an organism (e.g. a frog).

However, though the sorts of examples with which the literature has been concerned seem to bear this out they do not show that it is not *possible* for a representational system to have states which are causally sensitive to such properties. In general, my suggestion is that as long as the representational system is complex enough to have states which are causally sensitive to those properties, the characteristic effects of which qualify them for membership in some functional kind, it is not precluded, on a broadly causal account, from representing such properties.

And, undoubtedly there are such systems. For example, there is clear evidence that the pig regulates its control of food (both within a meal and over the course of meals) according to the nutritional content of the items ingested (Houpt, 1984).[98] One of the negative feedback mechanisms which is present in the pig to control dietary intake might be a glucostatic control (i.e. one based on the monitoring of metabolisable glucose in the digestive system) (Houpt, 1984). Supposing this sort of complexity to also exist in the frog, the situation is modelled in figure 9.

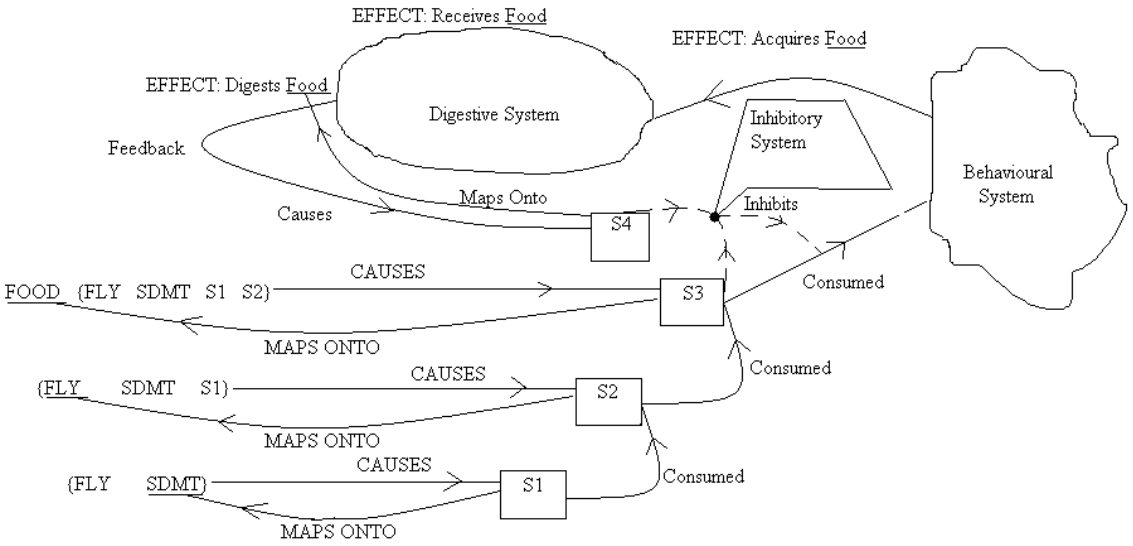

Figure 9

[98] For, example when the pig ingests items which have been mixed with indigestible material, food intake in subsequent meals is seen to increase (Houpt, 1984, p.1346).

Clearly, these sorts of controls are causally sensitive to exactly those properties in virtue of which an item qualifies as food (*qua* those properties). Moreover, supposing that state S4 in the diagram is an input into a device which modulates the effects of S3 on the behavioural system responsible for acquiring food, this device will only perform its Normal proper function when S4 maps onto food. Hence, by **(TRP$_{INF}$)** S4 has the content *food*.

This example is intended to demonstrate the possibility of some internal state being such that it is causally sensitive to the more distal concertina effects the mapping onto which is Normal. However, the general point is that as long as the properties with which we are concerned have certain (characteristic) effects, to which the system can be sensitive (in the appropriate sense) there is nothing in principle which precludes that system from representing those properties on an account of content which includes a (broadly) causal condition.

Similar points carry over to the case of having the concept of theoretical entities. However, in this case, since we are dealing with natural kinds, the effects must also be 'captured' in an intension. In general it seems that it is possible to have states which are sensitive to these characteristic effects, since it is most plausibly *by* such effects that we know about these properties, and we do take ourselves to have such knowledge. Even in the case of functional kinds, intensions may often play a significant role in the development of a state which is causally sensitive in the requisite way.

The simple cases with which the literature has often been preoccupied (e.g. frogs, toads, tortoises) have explored the causal production of a state such as S3 in isolation and consequently concluded that the more distal concertina effects play no role in the causal production of the state. The present proposal doesn't fall into that trap. However, the example, above, it seems doesn't give us quite what we want, since by **(TRP$_{INF}$)**, S3 does *not* have the content *food*. Yet the effects of S3 and S4 on the behavioural system seem to 'make sense' only if S3 has the same content as S4. Is this a problem?

I don't think so. We might say that **(TRP)** gives conditions on a content-like state (say 'proto-content', borrowing the term from Papineau (1993)), which helps to 'explain' events in the sense of showing historically how a certain subsystem/device/arrangement contributed toward fitness and was selected. However, over time, an organism may develop states which are causally sensitive to certain features of the environment which it proto-represents in proto-content. In our example, this would allow such states to have the content *food*. Such a state may come to be associated with the proto-content state (intuitively, in our example, by learning that the flies are food), and even to replace the latter state in terms of its effects on its consumers. This would be biologically advantageous, since it would allow the advantageous effects of producing certain behavioural responses in the presence of food to generalise beyond the organism's original environment.

The important point is that, given enough complexity (including complex movement patterns that require the organism to generalise across contexts), it is plausible to think that the processes appealed to in **(TRP_{INF})** could yield the genuine content *food* for S3. However, the fact that S3 needn't have this content shouldn't worry us excessively, since it is arguably only in such suitably complex systems that the possibility of rational explanation of behaviour becomes a requirement. Moreover, **(TRP_{INF})** seems to do a better job of such rational explanation. Indeed that **(TRP)** is insufficient in this regard is one of the points of Pietroski's (1992) cases; that **(TRP_{INF})** can accommodate them, therefore, counts in its favour.

Notice finally that on **(TRP_{INF})** representational structures would be selected in part for their ability to produce certain effects in the context of items to which they are causally sensitive. This is a *structural* feature of the state, whilst the effects produced determine the state's existence over time. Hence, the account accommodates Agar's points rather nicely.

6.1.2. ***VP***

In the discussion of **(TRP)**, the central problem raised for the account's response to **VP** was claimed to be the problem generated by the possibility of multiple consumers. Does **(TRP_{INF})** fare any better? For my own part I think that it does; however, it is difficult to give a decisive demonstration that this is the case. To see the central virtue of **(TRP_{INF})** with respect to this problem suppose that there are numerous representation consumers such that their Normal proper functioning does not render a unique item onto which the representation is to map. Still, it may be the case that not all of these items will fall under a type which bears an appropriate counterfactual supporting relation to the property of being a cause of representations of the type in question. Hence, **(TRP_{INF})** gives greater discriminative machinery and makes it far less likely that there will be genuine indeterminacy of content arising from the existence of multiple consumers.

Will there always be this singularity of content? Well arguably if there were, that would show that there is something wrong with the theory which generates the contents: For there are surely mental representations which are genuinely ambiguous, and, perhaps, disjunctive. What is required of a good theory of content is to show how it is possible for there to be univocal non-disjunctive contents, disjunctive contents and ambiguous contents (cf. Fodor, 1990b) and, I think, to allow or support the idea that the first set of contents is much larger than the second two. **(TRP_{INF})** does this quite nicely. A representational state will have a univocal determinate content if the mappings between the representation and the world which are Normal conditions for the proper functioning of each of the representation consumers converge onto one item mapped, and this item is such that the representational state is apt to carry information about it. A representation, **X**, will have a disjunctive content if there exists a representation consumer, C, there exist mappings, *M* and *M**, between **X** and **Y**, and between **X** and **Y***, respectively (**Y≠Y***), such that **X** is apt to carry information about **Y** and about **Y***, and each of *M* and *M** historically facilitated the successful performance of the proper functions of C as it responded to **X**.[99] Here the correspondence rules which partly

[99] Alternatively, though not equivalently, a representation, **X**, will have a disjunctive content if the mapping between the representation and the property picked out by a disjunctive predicate in English is a Normal condition for the performance of the proper function/s of the consumer of **X**.

constitute *M* and *M** may be the same. A representation, **X**, will have an ambiguous content if there exist representation consumers of **X**, C and C*, there exist mappings *M* and *M**, between **X** and **Y**, and between **X** and **Y***, respectively (**Y≠Y***), such that **X** is apt to carry information about **Y** and about **Y***, and *M* historically facilitated the successful performance of C's proper functions as C responded to **X**, and similarly with respect to *M** in the case of C*.[100] [101]

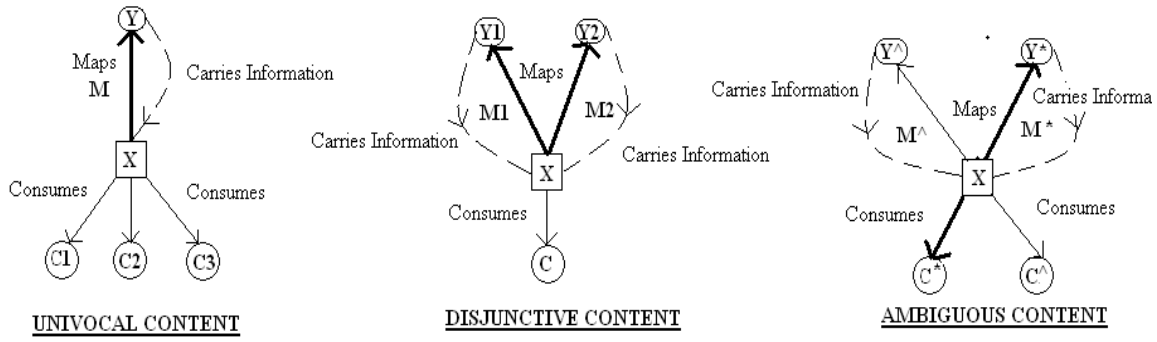

Figure 10

Since it seems plausible that in the case of paradigmatic representational systems, which will be complex, there are a number of representation consumers, and since **(TRP)** cannot give us principled reasons to suspect that the Normal mappings of such consumers will converge in the way required for definite content, **(TRP)** suffers from a theoretical disadvantage when compared to **(TRP_{INF})**. Of course, this latter account also cannot *guarantee* that there will always be univocality of content; and nor should it. However, it does give principled grounds

[100] We can generalize this definition of ambiguous content to the case in which there are more than two consumers. In this generalization, we require that the Normal mappings required by non-empty, disjoint subsets of the set of consumers converge onto different distinct items. Note, also that the definitions can interact with one another giving, e.g., ambiguous contents where the ambiguity is between two (or more) disjunctive contents.

[101] See figure 10 for a diagrammatic representation of these definitions.

to expect that univocality of content can be expected in the main; at least it gives us greater grounds for thinking this than does **(TRP)**.

More importantly it gives principled grounds for thinking that when there is not this sort of univocality, there is genuine ambiguity/disjunction of content, and it does this in a more satisfactory way than **(TRP)**. For, in the case where there is a plurality of content, neither the causal structure of the world as it pertains to the representational states of the organism, nor the needs/functioning of the organism as this pertains to the world can decide in favour of one determinate content over another. **(TRP_{INF}),** therefore, cannot decide between alternative contents (or render non-disjunctive contents) unless the world-organism and organism-world interactions can decide the matter. As far as a naturalistic theory of content goes, I think that this is the best that one can hope for; it is certainly better than relinquishing rights to univocal content just in case content cannot be univocally attributed in terms of one of these relationships, as **(TRP)** does.

6.1.3. ***PCR***

The lesson of the discussion thus far has been to suggest that an appeal to the cognitive complexity of a system can yield fruitful results. This is the crucial point which paves the way to an adequate resolution of **PCR**. The two main lessons which I will be taking from our discussion of this problem in the previous section are as follows:

(i) Neander's suggestion is promising, but unsuccessful when applied within a producer-based teleosemantic framework;

(ii) Papineau's focus on the degree of cognitive complexity of the organism is on the right track; however, the sort of complexity which is relevant is not the sophistication of the type of representational state, but rather the complexity of the system of states of which it forms a part;

The conclusion for which I wish to argue is that Millikan's suggestion can render contents which feature on both higher and lower levels of a concertina of proper functions when

suitably combined with revised versions of the suggestions of Neander (in line with (i)) and Papineau (in line with (ii)).

Consider figure 9, above, and figure 11, below.

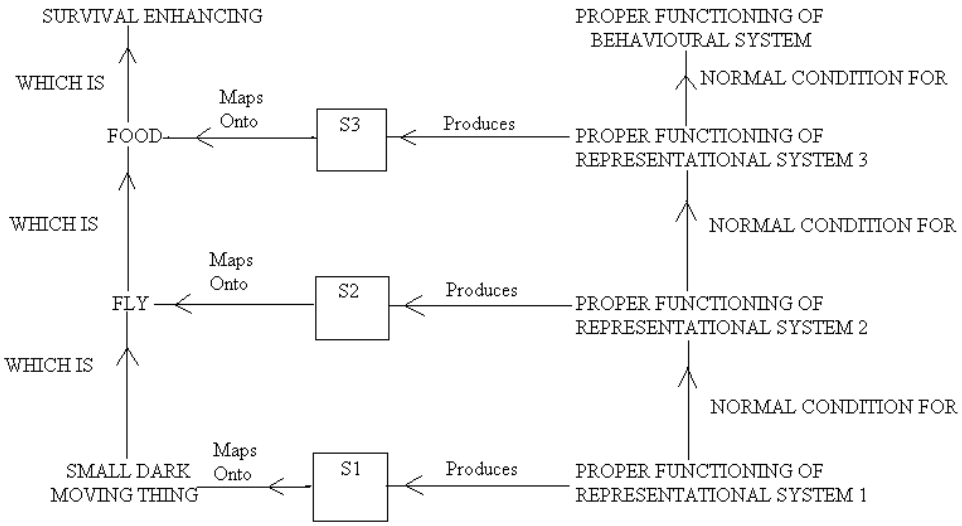

Figure 11

The digestive system has a concertina of proper functions of its own. However, by application of Neander's analysis of the notion of a specific function, this system has a function which is specific to it, namely the reception and digestion of food. The digestive system's specific function consequently determines, as the specific proper function of that part of the behavioural system which serves the digestive system, the acquisition of food: It is only when the behavioural system managed to acquire food that the digestive system historically performed its most specific proper function.

The basis for Neander's suggestion was the analysis of proper functions in biology; what caused the problem was her attempt to apply the same analysis directly to the representation producers within an organism. However, on a consumer-based semantics once we have applied the analysis to the representation *consumer*, no further application is necessary in

order to make the analysis relevant to representation, since it is the proper functioning of the consumer of a representation type which determines the content of that representation type.

Hence, the first criticism which was raised against Millikan's response to **PCR** (namely that either it cannot avoid assigning content from the topmost level of the concertina of functions or it must assign disjunctive contents) can be avoided. In the case of the frog, the mapping between representation, **S**, and the world which is a Normal condition for the successful performance of the *most specific* function of the representation consumer is the mapping between the representation and *food*. We have noted that the consumer focus of accounts such as **(TRP_{INF})** entails that the proper function of the representation producing system is to produce representations which map, *so*, onto the world. Hence, we might say that the most specific proper function of (that part of) the producing system which is responsible for **S** is to produce representations which map, *so*, onto the property *food*.

Now, suppose that the representation producing system is embedded within further representation producing systems, so that it is both a representation producer and a representation consumer.[102] The question is what mapping must obtain between S2 and the world, and between S1 and the world in order for representational systems 3 and 2 to successfully perform their Normal proper functions, respectively. The answer which I wish to give is that S2 must map, according to mapping M_2, onto *fly* and S1 must map, according to mapping M_1, onto *small, dark moving thing*. For, we have *assumed* that in the frog's Normal habitat food is Normally flies and flies are Normally small, dark and moving. Moreover, we have *assumed* that the frog succeeds in mapping onto ('detecting' in the original presentation) food *by* mapping onto flies, and that it succeeds in mapping onto flies by mapping onto small, dark moving things. Without this assumption the problem of **PCR** simply does not arise.

[102] The case is modelled in Figure 11, from which the designating terms which follow are taken, with 'representational system 3' corresponding to the producer system which is responsible for **S** (hence, **S** will be referred to as S3 from here on).

But notice that a Normal explanation is an explanation of *how* a proper function was historically performed on those (perhaps few) occasions when it was successfully performed (Millikan, 1989). That is which features of the environment were causally responsible for the successful performance of the relevant proper function. Hence, the problem of **PCR** arises only on the assumption that a concertina of means can be selected for achieving a concertina of (mostly intermediate) ends. However, *on this assumption* and on the assumption of the existence of suitable representational complexity in the relevant organism, it is correct to describe the Normal explanations for the performance of the proper functions of representational systems 1, 2 and 3, respectively, as including reference to the mappings between S1 and the world, S2 and the world and S3 and the world, respectively, which are represented in figures 8 and 10.

Notice that the *most proximate* Normal explanations will include reference *only* to the mapping at the level directly 'below' the system the proper performance of which the explanation explains. And it is only the most proximate explanation which is relevant to content determination in each case (cf. Millikan, 1989). Notice also that although each system in Figures 8 and 10 will have as proper functions all of the proper functions of the systems above it, applying Neander's analysis to each consumer respectively delimits only one representationally relevant function for each producer.

Hence, we are able to render determinate content to representational states within a teleological framework through the combination of certain central insights which can be extracted from the range of unsuccessful approaches to **PCR** in the literature. However, in order for this solution to generate determinate contents, the representational system must demonstrate a suitable complexity, in the respects indicated above, which matches that of the complexity of the environment as this environmental complexity pertains to the concertina of proper functions.[103]

[103] It should also be noted that this model of representation has something in common with (and perhaps gains some empirical support from the evidence informing) the computational models which Neander (1995) cites. For it conceives of representation of, e.g., medium-sized objects as being generated out of simpler representations of certain qualitative features of those objects. However, notice that this model does not depend upon inference in representation formation. Hence it avoids all of the difficulties which arose for Neander's

6.1.4. Counterintuitive Contents Again

One final issue should be addressed before closing. A further objection in the literature on accounts such as **(TRP)** has been that such accounts yield counterintuitive contents because they must render contents that are *overly* specific (Neander, 2004). Neander (2004) notes, for instance, that a variety of specific conditions can prevent a contribution to the fitness of the frog even when the frog ingests a fly/some food. For example if the fly is infected with some disease, or if a crow eats the frog soon after ingestion. Hence, she notes that it has been argued that an account such as **(TRP)** must render the content of the frog's representational state as *fly that has not been infected, when no crow is standing by...*, where the ellipses allow, in principle, for an indefinite number of further conditions to be added.

Notice, however, that **(TRP_{INF})** can easily avoid this consequence, when it is read as it has been suggested it should be in the previous subsection. In this form it has two points of recourse in answering the objection. The first is to point out that the *most specific* proper function of the representation consumers in the case of the frog is to acquire, to receive and to digest food; the most specific proper function is not to contribute toward fitness. It is only that mapping which is a Normal condition for the successful performance of the *most specific* proper function of the representation consumer/s which is determinative of content. Hence, even though there are numerous further conditions which must be met in order for the consumer to perform its *ultimate* proper function of contributing to fitness (and even though numerous factors external to the consumer will need to 'co-operate' in order for this to occur), this needn't affect the content attributions made to the organism. The second point of recourse is to claim that it is not true that such overly specific contents (*under the description which renders them overly specific*) cause tokens of internal representations (*qua* such representations).

---

account based on her suggestion that an appeal to inferential processes in perception could (or should) be made. Furthermore, the account does not require the simple representations to uniquely 'define' some medium sized object. This too is a good thing, because I doubt that any such definition is possible.

Hence, on either line of response, content is not necessarily overly specific. This is a virtue of the **(TRP$_{INF}$)**.

VII

CONCLUSION

Some have wondered whether we can "[b]ake a mental cake using only physical yeast and flour" (Dretske, 1981, p.xi). The above discussion has attempted to show that we can: Mix one part information with three parts teleology, throw in a handful of representation producers and consumers, add a cupful of isomorphic transformations and a jar of cognitive complexity, mix well and bake in a philosophical oven.

GLOSSARY OF ABBREVIATIONS

- **CR** = The Assumption of Content Realism: That representational properties are objective features of a certain class of states (i.e. mental states)
- **CE** = The Assumption of Content Externalism: That the content of (at least some) representational states is determined, at least in part, by the relations (past or present) such states bear to their external environment
- **PN** = The Assumption of Psychological Naturalism: That our best account of mental states and their representational properties should not posit or make appeal to entities or relations which are incompatible with (our most plausible version of) natural science
- **CQ** = The 'central question': In virtue of what sort of external, naturalistic relation does an internal, physical state come to either have or to realise a representational property.
- **ACA** = The assumption that there are mental representational states which express only properties and that there are such states which refer only to individuals
- **HP** = The Horizontal Problem: The problem of picking out that which is represented in a mental state from the other elements in the causal chain which eventually leads to the occurrence of an internal event.
- **VP** = The Vertical Problem: The problem of picking out the set of objects to which a mental state refers from, on the one hand, larger sets of which this set is a subset, and, on the other hand, smaller sets forming a subset of this set.
- **PIR** = The Problem of the Inscrutability of Reference: The problem of picking out the property expressed by a contentful state from other distinct yet necessarily co-instantiated properties.
- **DP** = The Problem of Disjunctivism: The problem of deflating the distinction between correct representation and misrepresentation by attributing a disjunctive content to a state.
- **PNC** = The Problem of Necessarily Co-referring Contents: The problem of distinguishing between representational contents which are intuitively distinct yet necessarily co-refer.
- **PER** = The Problem of Empty Representations: The problem of: (i) Accounting for the possibility of states which have content and no referent; (ii) accounting for how the content of such states could differ from one another.
- **I** = The Indication Relation: **X** indicates **Y** iff if an **X** occurs then a **Y** occurs.
- **DI_1** = The account of information found in Dretske (1981). Namely that: **X**(r) carries the information that **Y**(s) if and only if there exist some strict law of the form '□((...&**CC**)→...)' (where '**CC**' is a complete description of the elements of $\{C_i\}_{i\varepsilon[1,n]}$, under the relevant types) which is true for the substitution of **X** in the first open position and **Y** in the second, and there exists no law of the form '□(**CC**→...)' which is true for the substitution of **Y** in the second position.
- **DI_2** = The account of information found in Dretske (1986, 1988). Namely that **X**(r) carries the information that **Y**(s) if and only if P(**Y**,**X**) is sufficiently close to 1 and there is a contextually relevant condition which *explains* the existence/persistence of the correlation between **X** and **Y**.
- **FI** = Fodor's (1990a) account of information. Namely that: **X**s carry information about **Y**s (under that description) if and only if there is a nomic connection between the property of being a **Y** and the property of being a cause of **X**s. That is, the generalization "**Y**s cause **X**s" is a law (1990a, p.57).

- **IS** = **X** represents **Y** only if **X**s carry information about **Y**s (or alternatively, only if they do so under certain circumstances).
- **SC** = Dretske's (1981) account of the semantic content of a state type. Namely, for all r, s, for all **X**, **Y**, **X**(r) has **Y**(s) as its semantic content iff$_{def}$ **X**(r) carries the information that **Y**(s) in completely digitalized form.
- **MC$_{D1}$** = Dretske's (1981) account of mental content. Namely, for any **X**, **Y**, **X**(…)carries the information that **Y**(…) iff$_{def}$ during the learning period, L, instantiations of **X**(…) came to have **Y**(…) as their semantic content.
- **LP** = The reconstruction of Dretske's (1981) notion of the learning period. Namely: A period of time [t_1, t_n] constitutes a learning period for a mental representation, **X**(r) iff there exists some state of affairs **F**(s) such that:
  (i) At t_1 **X**(r) codes in analog form the information that **F**(s), relative to channel conditions **CC**;
  (ii) At t_n **X**(r) codes, for the first time, in completely digitalised form the information that **F**(s), relative to channel conditions **CC**.
- **L** = The learning period in Dretske (1981).
- **AD** = Fodor's asymmetric dependence account of mental representation. Namely, For any representational state type, **X**, for any property, **Y**, tokens of **X** express (represent) **Y** if:

(iii) There is a nomic dependence between the property of being a **Y** and the property of being a cause of **X**s, such that the generalisation '**Y**s cause **X**s' is lawlike.

(iv) For any property **Z**, if **Z**≠**Y** and there is a nomic dependence between the property of being a **Z** and the property of being a cause of **X**s then the law that **Z**s cause **X**s is asymmetrically dependent on the law that **Y**s cause **X**s.

- **AC** = Fodor's asymmetric dependence condition. Namely, for any properties **X**, **Y**, **Z**, if **Z**≠**Y** and '**Y**→**X**' is a law and '**Z**→**X**' is a law, then the nomic dependence between **Z** and **X** asymmetrically depends on the nomic dependence between **Y** and **X** if and only if:

(iii) If '**Y**→**X**' were to be broken, then '**Z**→**X**' would be broken;

(iv) If '**Z**→**X**' were to be broken then '**Y**→**X**' would still hold.

- **ACP** = The interpretation of **AC** in the language of possible worlds semantics. Namely, for any properties **X**, **Y**, **Z**, if **Z**≠**Y** and '**Y**→**X**' is a law and '**Z**→**X**' is a law, then the nomic dependence between **Z** and **X** asymmetrically depends on the nomic dependence between **Y** and **X** if and only if:

(i) There exists some possible world, W, such that in W '**Y**s cause **X**s' is a law but '**Z**s cause **X**s' is not;

(ii) For every world, W*, if in W* '**Z**s cause **X**s' is a law but '**Y**s cause **X**s' is not, then W is closer to @ than is W*.

- **TIS** = Teleological Indicator Semantics: The thesis that an inner state type **X** represents a type of state of affairs, **Y**, iff **X** has the function of indicating **Y**.
- **TRC** = Representation Consumer Focused Teleological Semantics: The thesis that an inner state **X** is a representation and a representation of some condition **Y** iff: (a) that

X correspond by a certain rule with **Y** is a necessary condition on the performance of the proper function, in accordance with a Normal explanation, of the consumers of **X**; (b) the class of representeds, into which **Y** falls, admits of variations which accord with the conjunction of the set of correspondence rules and the rules defining the semantics of the system of representations.

- **PCR** = The Problem of Complex Causal Roles: The problem that biological proper functions are not/cannot be determinate in the way required by teleological theories of mental content because an effect's causal contribution to the selection of the trait which is responsible for it is realized by a number of intermediary mechanisms.
- **NPF** = Neander's interpretation of the notion of malfunction. Namely, that a subsystem will be considered to be malfunctioning when it fails to produce its biologically normal effect at the lowest level of description/analysis at which it appears explicitly and as an unanalysed component.

- **NPM** = For any system, **Q**, responsible for producing a representational state, **S**, **Q**'s tokening of **S** is a misrepresentation if and only if **Q** malfunctions (according to **NPF**) in producing **S**.

- **NPF*** = The revision of **NPF** which attempts to accommodate distal contents. Namely: A subsystem is malfunctioning if it fails to successfully perform the lowest of its biologically normal functions *within some class*, $\boldsymbol{E_D}$. Here, $\boldsymbol{E_D}$ is: (i) a subset of the concertina of functions generated by the description of the subsystem when it appears explicitly and as an unanalysed component; (ii) determined by certain further principles to which Neander appeals.

- **NPM*** = The revision of **NPM** which attempts to accommodate distal contents. Namely: For any system, **Q**, responsible for producing a representational state, **S**, **Q**'s tokening of **S** is a misrepresentation if and only if **Q** malfunctions (according to **NPF***) in producing **S**.

- $\mathbf{IC_N}$ = The account of information adopted in section IV. Namely: For all **X**, for all **Y**, **X**s carry the information that **Y** if and only if:

  (i) **Y**s *qua* **Y**s cause **X**s *qua* **X**s;

  (ii) **Y**s have actually caused **X**s;

  (iii) $0<P(\mathbf{Y},\mathbf{X})$.

- $\mathbf{TRP_{INF}}$ = The account of mental content/representation adopted in section VI. Namely: An inner state **X** is a representation of some condition **Y** iff:

  (i) **X**s are apt for carrying the information that **Y**;

  (ii) that **X** is related by a certain correspondence rule, *R*, to **Y** is a necessary condition on the performance of the Normal proper function of the consumers of **X**;

(iii) **Y** admits of variations, within the class of representeds into which it falls, which accord, one-to-one, with a set of rule-governed transformations of the class of representations into which **X** falls;

(iv) the rule-governed transformations of **X** have as a Normal condition of their proper function the corresponding variation of **Y**.

BIBLIOGRAPHY

Agar, N. (1993). What Do Frogs Really Believe. *Australasian Journal of Philosophy , 71*, 1-12.

Antony, L. & Levine, J. (1991). The Nomic and the Robust. In B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and his Critics* (pp.1-16). Cambridge, MA: Blackwell.

Armstrong, D. (1980). The Causal Theory of Mind. In D. Armstrong, *The Nature of Mind* (pp. 16-31). St Lucia, Queensland: University of Queensland Press.

Baker, L.R. (1991). Has Content Been Naturalised? In B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and his Critics* (pp.17-32). Cambridge, MA: Blackwell.

Boghossian, P.A. (1991). Naturalising Content. In B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and his Critics* (pp.65-86). Cambridge, MA: Blackwell.

Braddon-Mitchell, D., & Jackson, F. (1997). The Teleological Theory of Content. *Australasian Journal of Philosophy , 75*, 474-89.

Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy, 4*, 73-121.

Burge, T. (1986). Individualism and Psychology. *Philosophical Review, 95*, 3-45.

Chalmers, D. (2006). Two-Dimensional Semantics. In E. Lepore, & B. Smith (eds), *The Oxford Handbook of the Philosophy of Language* (pp. 574-606). New York: Oxford University Press.

Cram, H. (1992). Fodor's Causal Theory of Representation. *The Philosophical Quarterly, 42*, 56-70.

Davidson, D. (1986). A Coherence Theory of Truth and Knowledge. In E. Lepore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (pp.307-319). Oxford: Basil Blackwell.

Davidson, D. (1967). Causal Relations. In E. Sosa, *Causation and Conditionals* (pp. 82-94). Oxford: Oxford University Press.

Davidson, D. (1987). Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association , 60* (3), 441-58.

Dennett, D. (1987). Evolution, Error and Intentionality. In D. Dennett, *The Intentional Stance* (287-321). Cambridge, MA: MIT Press.

Devitt, M., & Sterelny, K. (1999). *Language and Reality: An Introduction to the Philosophy of Language* (2nd Edition ed.). Cambridge , MA: The MIT Press.

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

Dretske, F. (1983). The Epistemology of Belief. *Synthese , 55* (1), 3-19.

Dretske, F. (1986). Misrepresentation. In R. Bogdan (ed.), *Belief: Form, Content and Function* (pp. 17-36). New York: Oxford University Press.

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

Field, H. (1978). Mental Representation. *Erkenntnis , 13* (1), 9-61.

Fodor, J. (1984). Semantics, Wisconsin Style. *Synthese , 59* (3), 231-250.

Fodor, J. (1987). *Psychosemantics.* Cambridge, MA: Bradford Books.

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

Fodor, J. (1990a). A Theory of Content I: The Problem. In J. Fodor, *A Theory of Content and Other Essays* (pp.51-88). Cambridge, MA: Bradford Books/MIT Press.

Fodor, J. (1990b). A Theory of Content II: The Theory. In J. Fodor, *A Theory of Content and Other Essays* (pp.89-136). Cambridge, MA: Bradford Books/MIT Press.

Fodor, J. (1991). Replies. In B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and his Critics* (pp.255-319). Cambridge, MA: Blackwell.

Fodor, J. (1994). *The Elm and the Expert: Mentalese and its Semantics*. Cambridge, MA: MIT Press.

Fodor, J. (2008). Against Darwinism. *Mind and Language , 23,* 1-24.

Gates, G. (1996). The Price of Information. *Synthese , 107* (3), 325-347.

Gende, S.M., Quinn, T.P. & Wilson, M.F. (2001). Consumption Choice by Bears Feeding on Salmon. *Oecologia 127*(3), pp.372-382.

Godfrey-Smith, P. (1989). Misinformation. *Canadian Journal of Philosophy , 19* (4), 533-50.

Godfrey-Smith, P. (1992). Indication and Adaptation. *Synthese , 92* (2), 283-312.

Godfrey-Smith, P. (1996). Meaning, Models and Selection: A Review of Philosophical Naturalism. *Philosophy and Phenomenological Research , 56* (3), 673-78.

Godfrey-Smith. (2002). Environmental Complexity and the Evolution of Cognition. In R. Sternberg, & J. Kaufman (eds), *The Evolution of Intelligence* (pp. 233-249). Mahwah: Lawrence Erlbaum.

Godfrey-Smith, P. (2004). On Folk Psychology and Mental Representation. In H. Clapin, P. Staines, & P. Slezak, *Representation in Mind: New Approaches to Mental Representation* (pp. 147-162). Amsterdam: Elsevier.

Gould, S., & Lewontin, R. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences , 205* (1161), 581-98.

Greenberg, M., & Harman, G. (2005). Conceptual Role Semantics. In E. Lepore, & B. Smith (eds), *Oxford Handbook of the Philosophy of Language.* UCLA School of Law Research Paper No 05-24.Available at SSRN:http://ssrn.com/abstract=800187: Oxford University Press.

Griffiths, P., & Goode, P. (1995). The Misuse of Sober's Selection for/Selection of Distinction. *Biology and Philosophy , 10*, 99-107.

Hogg, R.V. & Tanis, E.A. (2010). *Probability and Statistical Inference*. New Jersey: Pearson/Prentice-Hall.

Horgan, T. (1989). Mental Quasation. *Philosophical Perspectives, 3*, 47-74.

Horgan, T., & Graham, G. (1991). In Defense of Southern Fundamentalism. *Philosophical Studies , 62* (2), 107-134.

Houpt, R. (1984). Controls of Feeding in the Pig. *Journal of Animal Science 59*(5), pp.1345-1353.

Kelly, T. (Fall 2008). Evidence. In E. Zalta (ed.), *Stanford Encyclopaedia of Philosophy*. URL =<http://plato.stanford.edu/archives/fall2008/entries/evidence/>.

Kripke, S. (1980). *Naming and Necessity.* Cambridge, MA: Harvard University Press.

Leon, M. (1987). Solipsism Regained. *Analysis , 47* (2), 116-120.

Leon, M. (1988). Realism, Skeptism and Empiricism. *Metaphilosophy , 19* (2), 143-157.

Leon, M. (1990). The Mechanics of Rationality. *The Southern Journal of Philosophy , 27* (3), 343-366.

Leon, M. (1992). Rationalising Belief. *Philosophical Papers , 21* (3), 299-314.

Leon, M. (1996). Sensations, Error and Eliminative Materialism. *The Southern Journal of Philosophy, 34* (1), 83-95.

Lewis, D. (1980). Mad Pain and Martian Pain. In N. Block (ed.), *Readings in the Philosophy of Psychology* (Vol. 1, pp. 216-22). Cambridge, MA: Harvard University Press.

Lloyd, D. (1987). Mental Representation from the Bottom Up. *Synthese , 70* (1), 23-78.

Loewer, B. (1987). From information to Intentionality. *Synthese, 70*, 287-317.

Loewer, B. (2002). *A Guide to Naturalizing Semantics.* Retrieved March 2011, from http://www.nyu.edu/gsas/dept/philo/courses/rules/papers/Loewer.pdf

McGinn, C. (1979). Action and its Explanation. In N. Bolton (ed.), *Philosophical Problems in Psychology*. London: Methuen.

McGinn, C. (1982). The Structure of Content. In C. McGinn, *Knowledge and Reality, Selected Essays* (pp. 111-151). Oxford: Oxford University Press.

McLaughlin, B.P. (1991). Belief Individuation. In B.P. McLaughlin (ed.), *Dretske and His Critics* (pp.157-179). Cambridge, MA: Blackwell.

Millikan, R. (1984). *Language, Thought and Other Biological Categories*. Cambridge, MA: Bradford Books/MIT Press.

Millikan, R. (1986). Thoughts Without Laws: Cognitive Science with Content. *The Philosophical Review , 95* (1), 47-80.

Millikan, R. (1989). Biosemantics. *The Journal of Philosophy , 86* (July), 281-97.

Millikan, R. (1990). Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox. *The Philosophical Review , 99* (3), 323-353.

Millikan, R. (1991). Speaking Up for Darwin. In B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and his Critics* (pp.151-164). Cambridge, MA: Blackwell.

Millikan, R. (2000). *On Clear and Confused Ideas: An Essay on Substance Concepts*. Cambridge: Cambridge University Press.

Millikan, R. (2004). *Varieties of Meaning*. Cambridge, MA: MIT Press.

Millikan, R. (2007). An Input Condition for Teleosemantics? Reply to Shea (and Godfrey-Smith). *Philosophy and Phenomenological Research, LXXV*(2), pp. 436-455.

Neander, K. (1995). Malfunctioning and Misrepresenting. *Philosophical Studies, 79*, 109-41.

Neander, K. (2004). Teleological Theories of Mental Content. In E. Zalta (ed.), *The Stanford Encyclopaedia of Philosophy (Winter 2009 Edition).* URL = <http://plato.stanfor.edu/archives/win2009/entries/content-teleological/>.

Nicholson, A. (2009). The Problem of Syntactitude. Unpublished manuscript.

Nicholson, A. (2012). Dretske and Fodor on Information. Unpublished manuscript.

Papineau, D. (1993). *Philosophical Naturalism.* Oxford: Blackwell.

Papineau, D. (1996). Reply to Commentators. *Philosophy and Phenomenological Research , 56* (3), 687-97.

Papineau, D. (1997). Teleosemantics and Indeterminacy. *Australasian Journal of Philosophy, 76*, 1-14.

Papineau, D. (2001). The Status of Teleosemantics, or How to Stop Worrying about Swampman. *Australasian Journal of Philosophy , 79* (2), 279-89.

Pinker, S. (1995). *The Language Instinct*. London: Penguin.

Putnam, H. (1975). The Meaning of 'Meaning'. In K. Gunderson, *Language, Mind and Knowledge, Minnesota Studies in the Philosophy of Science, vii*, (pp. 131-193). Minneapolis: University of Minnesota Press.

Putnam, H. (1981). *Reason, Truth and History.* Cambridge: Cambridge University Press.

Quine, W. (1959). Meaning and Translation. In R. Brower (ed.), *On Translation* (pp.148-172). Cambridge, MA: Harvard University Press.

Shani, I. (2007). Intension and Representation: Quine's Indeterminacy Thesis Revisited. *Philosophical Psychology, 18*(4), 415-440.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal, 27*, July, October, pp. 379-423, 623-656.

Shea, N. (2010). Millikan's Isomorphism Requirement. Forthcoming in J. Kingsbury and D. Ryder (eds), *Millikan and her Critics*. Oxford: Blackwell. Draft of 14 September, available at: http://www.philosophy.ox.ac.uk/_data/asses/pdf_file/0014/25340/Shea,_Millikans_Isom_reqt,_revised_dr,_14Sep10.pdf.

Sober, E. (2010). Natural Selection, Causality, and Laws: What Fodor and Piatelli-Palmarini Got Wrong. *Philosophy of Science , 77*, 594-607.

Stampe, D. (1977). Toward a Causal Theory of Linguistic Representation. *Midwest Studies in Philosophy: Studies in the Philosophy of Language, 2*, 42-63. Minneapolis: University of Minnesota Press.

Sterelny, K. (1990). *The Representational Theory of Mind: An Introduction.* Cambridge, MA: Blackwell.

Stich, S. (1990). *The Fragmentation of Reason.* Cambridge, MA: The MIT Press.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.