MANOLO MARTÍNEZ

A NATURALISTIC ACCOUNT OF CONTENT AND AN APPLICATION TO MODAL EPISTEMOLOGY

Manolo Martínez: *A Naturalistic Account of Content and an Application to Modal Epistemology,* abril de 2010

PROGRAMA:
Cognitive Cience and Language PhD Program

SUPERVISOR:
David Pineda

TUTOR:
Manuel García-Carpintero

DEPARTAMENTO:
Department de Lògica, Història i Filosofia de la Ciència

FACULTAD:
Facultat de Filosofia

UNIVERSIDAD:
Universitat de Barcelona

BIENIO:
2004-2006

A Blanca y Mateo.

A mis padres y mi hermano.

A los de Logos.

ABSTRACT

A naturalistic account of mental content is presented, aimed at solving the problems with content indeterminacy that afflict other similar efforts -notably, teleosemantics. Along the way, the sketch of a naturalistic account of compositionality in thought is also provided.

This account of content developed in the first part is then used to provide the foundations for a naturalistic modal epistemology.

RESUMEN

Se presenta una teoría naturalista acerca del contenido mental que pretende resolver los problemas de indeterminación del contenido a los que se enfrentan otras teorías similares, la teleosemántica entre ellas. Como parte de esta presentación, se ofrece el esbozo de una teoría naturalista acerca de la composicionalidad del pensamiento.

La teoría acerca del contenido desarrollada en la primera parte se usa a continuación como fundamento de una epistemología modal naturalista.

ACKNOWLEDGEMENTS

It is very easy, and very gratifying, to do philosophy in the environment created by the Logos group in the philosophy department of the University of Barcelona. Many, many mistakes in this dissertation have been avoided thanks to the constant help and criticism of Logos members, and many more would surely have been avoided had I payed more attention. This is why this dissertation is also dedicated to them -not just to the ones based in the UB, but also to those at the Autònoma, Girona, Argentina, Canada or Scotland.

Among Logosians I wish to acknowledge, in particular, the help of my supervisor, David Pineda, and of those other members that have read and commented substantial portions of the dissertation: Marc Artiga, Jose Díez, Manolo García-Carpintero, Genoveva Martí and Miguel Ángel Sebastián among others.

I have also profited from discussion with audiences in Buenos Aires, Santiago de Compostela, Lausanne, Amsterdam, Tarragona and Barcelona, and with David Buller, Jerry Fodor, Nicholas Shea and Michael Tye. Ruth Millikan sent me a good number of pages of comments to a draft of chapter 1, after apologising (!) for not doing it quickly enough. I had heard that she is very helpful to graduate students, but that's an understatement.

I would also like to thank the online community that has made the actual process of writing and typesetting this dissertation so much easier, by developing the software and helping newbies like me to undertand and benefit from them. Among many others: the LyX developers that read the LyX-users mailing list, André Miede for creating the beautiful LaTeX syle I have used to typeset the dissertation, and Nick Mariette who ported it to LyX. Details on how to get the amazing set of tools provided by this group of incredibly generous people can be found in page 229.

My brother Pablo has drawn Arete and her family for chapters 1 and 2. I am one of the few people that remembers that he can draw; I think he'd like me to forget.

My friends have consistently pressed the "I like it" button in my status updates.

Blanca is my favourite person, and I love her very much.

The initial goal of this dissertation was to make a contribution to the naturalisation of modal epistemology. I had the idea of using a broadly teleosemantic notion of concept to explain that our concevings are reliably connected with the modal realm: concepts arise out of our causal interaction with worldly entities, and it might well be that those interactions allow them to encode some modal properties of these entities. This encoding would underpin the reliability of our conceivings.

In the actual course of my research, though, I have found issues that have compelled me to linger in the discussion of the (at first, supposed to be merely introductory) remarks on naturalistic accounts of content. So much so, that the actual dissertation I am submitting is mostly, in its longer first part, a presentation of such an account of mental content; while the discussion of modal contents and modal epistemology has ended up as an application and extension of this account, in the shorter second part.

In particular, I was dissatisfied with several features of the best worked-out naturalistic theories of content -and, among them, Millikan's biosemantics, which I take to be the best of all. First and foremost, these accounts are, to a higher or lesser degree, subject to an indeterminacy problem: they are not able to make univocal predictions as regards the content of mental states. This problem has been around for at least twenty years, maybe longer and, while it was widely discussed some ten years ago, it has largely fallen out of fashion, with next to no new contributions as of late. As is painfully often the case in philosophy, this has not happened because researchers feel that the issue has been solved but, rather, because they sense that everybody has said what they had to say, and a stalemate has been reached. Chapters 1 and 2 are an attempt at unblocking this situation.

I have the feeling that a consensus is emerging, among those still interested in the problem, according to which the right solution to the conundrum of the naturalisation of content will have to draw from ideas coming from both producer and consumer teleosemantics -that the right theory will be a *hybrid theory*. Shea (2007)'s *infotel semantics*, Ryder (2004)'s *SINBAD neurosemantics* and maybe Rupert (2008)'s *causal-developmental theory* are part of this growing consensus; Millikan too has seen it useful to spell out, in her (2004), how her theory explains that contentful states carry information about their content. I also believe that the right theory must be hybrid: the simplest contentful states exploit a correlation between detectable properties (*Being shiny and black*, say) and useful properties (*e. g., Being nutritious*, or *Being dangerous*). If we start from the useful-properties end, and try to recover the detectable-properties end, we encounter the Output Problem: our content attributions to, *e. g.*, animals that seek to mate will, implausibly, involve things like *Fertile mate that will not die before reproducing*. If, on the other hand, we start from the detectable-properties end, we face the Input Problem: our content attributions will involve implausibly proximal properties.

We need to start from both ends at the same time. And maybe there are not just two ends, maybe usefulness and detectability are not the only interesting dimensions along which properties can vary for the purposes of the content naturaliser. We need to start from *all* ends at the same time: contents involve structures formed by the coinstantiation of a mix of detectable and useful properties, and not only those, together with the explanation that they are so coinstantiated. The account of content defended in this dissertation is a development of these ideas.

A second issue in the theory of mental content, at least if you buy a teleosemantic approach and an etiological theory of functions, is how to account for the existence of contentful mental states that have not been selected. This is of the utmost importance, if only because *all* of the mental states we are interested in are of this kind. In this case, I take the main proposal available to be largely correct: not selected mental states have content because they have been produced by selected mechanisms. On the other hand, I'm not as convinced of the actual implementation details: the main idea available is that products of selected mechanisms also have a function in virtue of the fact of being such products. But I am not sure that products have this normative dimension. For instance, it does not seem right to say that a token bee dance that leads other bees down a path without honey *should* have done otherwise; this is true for the bee-dance producer, of course, but where does the token dance itself get its "should" from? I don't think that a principled answer to this question has been provided and, anyway, I think that content attributions to ephemeral states can be sustained making appeal only to the function of the selected producer. This involves extensive use of the *structures* I was alluding to in the previous paragraph, and I think this is as it should be: contents are tools for dealing with real world entities that exploit connections among those entities. All of this must be in place, out there, for contents to earn their keep in our cognitive setup.

The third issue I tackle in the dissertation -and the second in importance, after indeterminacy- is that of productivity and compositionality. This is not a case in which I think the standing theories are wrong, or otherwise unsuccesful; in this case, the problem has been one of insufficient attention: apart from Millikan's theory of mapping functions -and even this is sketchier than other areas of her theory- there are no substantial proposals on how to account for the compositionality of thought, starting from a broadly teleosemantic perspective. I can think of two reasons why this is so. First, there are sufficiently important problems in the foundations of teleosemantics for researchers to feel that it is premature to worry about compositionality -for example, one does not need structure in thought to pose the indeterminacy problem. The second reason is, maybe, the implicit assumption on the part of some researchers that a teleosemantic recipe for the content of concepts (such as HORSE) may be given, and that the account of compositionality could then be simply co-opted from that offered by the computational theory of mind (CTM), or some variation thereof. I also believe it would be nice to think of compositionality along the lines offered by CTM, but I am aware of the need to start from propositional contents as the most basic ones -that is to say, I believe that at some stage in the phylogenesis of contents there are only propositional contents, with concepts a later innovation. As the chapters leading to (and including) chapter 4 will show, there is no other way to make teleosemantics (and related

approaches, such as mine) work. Chapter 4 may be, then, seen as a proposal on how to make compatible the kind of naturalistic account of content I favour with the CTM take on compositionality. The discussion of compositionality in thought closes the first part of the dissertation, dedicated to the extensive spring cleaning of teleosemantics I have just summarised.

Apart from the need to solve these three problems with naturalistic accounts of content, there was another constraint informing the discussion: I wanted my theory to be suitable for extension to the modal case. While there has been a recent surge of interest in the naturalisation of modal epistemology, there has been virtually no overlap of the work of naturalistic modal epistemologists and that of content naturalisers. This is an awkward situation: surely the people that have devoted their writings to uncovering the nature of conceptual content should have something to say about, for instance, how we are able to entertain thoughts regarding merely possible states of affairs? In fact, the very few references to this family of problems that I have been able to find in the writings of content naturalisers show, more than indifference, distrust. I have formed the impression that people such as Millikan regard modality as one of these areas in which substantial progress is not likely to be forthcoming. Phenomenal consciousness being, maybe, another of these areas.

The second part of my dissertation may, then, be regarded as community service: it will have have fulfilled an important goal if it convinces other philosophers of the possibility of achieving progress in the naturalisation of modal content, and of the convenience of starting out from naturalistic accounts of content in general in so doing. Indeed, most of what I say in the second part falls out from, or are otherwise straightforward extensions of, what I say in the first part: contents involving the merely possible are entertainable because the open future is thinkable -this I show using the very same tools that show that *There is a fly around* is thinkable-, and many (maybe all) thoughts involving possibilities are simply thoughts about what is or has been an open future possibility. The last chapter of this second part discusses how an account of content along the lines developed here can be fruitfully used to inform a naturalistic account of epistemology so as to solve the well-known Generality Problem. In particular, I sketch how an account of modal content can inform modal epistemology.

I would like to raise a caveat regarding my interpretation of Millikan. Although I have made every effort to portray her ideas faithfully, it is entirely possible that I have occasionally misrepresented them. In particular, it is possible that I sometimes represent myself as in disagreement with her when, in fact, I am simply adhering to her opinions. If there are such cases, I hope that the unwitting reexposition of Millikan's theses sheds some light on them, and provides elaboration to points she has left in a sketchier level of detail.

I spend a substantial portion of this dissertation disagreeing with Millikan, and I feel obliged to say what will surely be obvious to every reader of the pages to follow: the disagreement takes place against a background of general agreement.

In this dissertation, definitions are introduced and quoted using SPACED LOW CAPS. An index of frequently used definitions may be found in page 227.

Mental mechanisms's names are written in LOW CAPS. Many times, 'N' stands for a mechanism that produces other mechanisms such as M.

The content of thoughts -such as *There is a fly around* is written in *italics*; also the name of properties such as *Being red*. Sections within this work are quoted thus: 1.3.1. In the pdf version that was a hyperlink.

CONTENTS

LIST OF FIGURES

Part I

CONTENT

1 A SOLUTION TO THE INDETERMINACY PROBLEM

One of the most vexing problems for naturalistic accounts of content is that of indeterminacy. In a higher or lesser degree, all of the proposals on offer in the marketplace for the naturalisation of the content of mental states are such that they can be shown to yield indeterminate results about what exactly these contents are. Solving this problem is, therefore, a main stepping stone in the road to a fully working naturalistic semantics.

In this first chapter I defend a solution, which may be seen as a refinement of teleosemantics, to this Indeterminacy Problem for simple contentful states. The idea, in a nutshell, is the following: in the vast majority of cases, selection of an indicator M is made possible by the presence of certain natural structures that may be identified with Homeostatic Property Clusters (HPC). Regardless of the different possible descriptions of the facts that play an explanatory role in the existence of M, and regardless of the different possible function-attributions that such explanations ground, content attributions to M's being *on* must involve said HPCs.

After a quick review of some of the literature on causal accounts of content, in section 1.2 I introduce a broadly Dretskean naturalistic account of content, for a simple mechanism M such that the content of its positives is of the form *There is an F around*. In the formulation of such an account, two notions are used (*indication* and *function*) that may raise suspicions as regards their naturalistic credentials. I quickly review a widely accepted way of analysing them in clearly non-intentional terms.

In section 1.3 I introduce the Indeterminacy Problem: M's fitness contribution -which is the one property of M's that natural selection ultimately favours- may be seen from multiple perspectives. From each perspective, the explanatorily relevant fact is that M indicates the instantiation of property F_i, for an indefinitely high number i of properties. Each of these indication relations, indeed, grounds a content-attribution and there is no principled reason to choose among all of them.

Fortunately, there are more facts among those that explain the existence of M than the ones used for fixing M's etiological function. Thus, in section 1.4 I will argue that such explanation may invoke the causal groundings of the following fact: a mechanism with M's causal powers remains a positive contributor to the fitness of its possessor *across generations*. These causal groundings may, and in the vast majority of cases will, include a homeostatic mechanism in the world that explains the recurrence of the properties used in the explanation of M's fitness contribution in each particular generation. Such a homeostatic mechanism, together with the properties the recurrence of which it explains, fix a Homeostatic Property Cluster. My proposal is that it is *these* HPCs that must be used in the content attribution for the positives of M. In section 1.5 I show how this proposal deals with the Indeterminacy Problem, and in section 1.6 I take up the objection that my proposal (which I call *Etiosemantics* to emphasise the role that causal explanations of

the existence of representations play in fixing their content) is not a proposal about *content*.

The preliminary conclusion of the chapter is that etiosemantics provides a recipe for the attribution of content to simple, selected-for mental states that is both intuitive and immune to the problems that vex most other naturalistic accounts.

1.1 NATURALISING CONTENTS

Some entities are *about* others, or *refer to* or *involve* others. For example, many sentences in my copy of Dummett's *Frege* are about Frege. My belief that the Earth is a planet refers to the state of affairs that consists of the Earth's being a planet. My current perceptual state involves the screen of my laptop. All of these entities that are about, refer to and involve others are said to have *content*. It is in virtue of having the content they have that they so refer, involve and display aboutness.

Contents are non-negotiable tools in a wide range of theoretical pursuits. Everyone everyday relies on some of them. Take belief and desire attributions: the behaviour predicting strategy that postulates the existence of these paradigmatically contentful mental states does not seem the kind of thing we may be ready to abandon anytime soon.

Some other of these pursuits are of interest only to philosophers. Such is, for example, the project of characterising mentality; the idea (popular at least since Brentano) being that intentionality (*i. e.*, aboutness, the having of content) is the mark of the mental. Another use philosophers put contents to -and about which I will have something to say in later chapters- is in the theory of how we get to know that some state of affairs is necessary, or that some thought could have referred to an actual state of affairs -although, as it happens, it does not. A popular idea is that our knowledge of the necessary and the possible is mediated by a particular kind of attitude to contents we may call *finding conceivable*. What we find or fail to find conceivable are, again, contents.

On the other hand, contents, such most useful entities, have a perplexing feature. We do not seem to have a clear answer to the question: in virtue of *what* do contentful entities have them? One not very satisfactory answer is that having a content is a brute characteristic of the mental (and mind-related entities, such as linguistic outputs), not reducible to or otherwise explicable in terms of other, more basic entities. This is not very satisfactory because, given the sweeping success of scientific explanation, it seems more and more plausible that the only brute, non reducible properties are physical properties. That content/physical-stuff dualism is true seems, thus, extremely unlikely. In any event, if a fully naturalistic account of content can be worked out, it is to be preferred if only on grounds of simplicity.

Fortunately, there are several promising approaches to the naturalisation of content. The theory I will be defending in this Part I is an example of what have been called *causal-informational accounts* of content, so in this introduction I will be paying attention to other examples of this approach[1]. In other sections of this and the next chapter I will discuss in some more depth other examples of this approach, more germane to my own proposal, such as Dretske's theory of content

1 There are many reviews of the literature on content naturalisation. In preparing this summary I have found useful Fodor (1990), Loewer (1999), Papineau (2006) and Rupert (2008).

circa Dretske 1988, Millikan's *biosemantics*, Shea's *infotel-semantics* and Papineau's version of teleosemantics.

Besides causal-informational accounts there is another important tradition in the naturalisation of contents, which offers what we may call *network-based accounts*. Causal and Inferential Role Semantics, in their various flavours, are important network-based accounts of content. Although I will largely ignore this other approach, I will have something to say about it in section 4.8.

1.1.1 *Very Simple Causal Accounts*

As a starting point in the project of naturalising content we may ask: which physical relation between representation and represented provides the best raw material for the construction of intentionality? One possibility is geometric/structural similarity[2]. For example, one could defend the following simple theory about the content of a perception of my mother's face:

STRUCTURAL SIMILARITY: A mental state MMF is an Idea of my mother's face if and only if it bears the relevant kind of structural similarity with her face.

This kind of structure-based account is seldom defended these days. Even granting that one can spell out which is the relevant kind of structural similarity involved, one main difficulty with this account comes from appreciating that resemblance is symmetrical, while intentional relations are not: my Idea of my mother's face is about her face, but her face is not about my Idea -*cf.* Wittgenstein (1953/1973). Another source of objections is the work in externalist semantics, from the seminal Kripke (1980) and Putnam (1975) on. This work has convinced many philosophers that no amount of, say, geometrical or structural information encoded in our mental state can make it be about *my mother's face* -as opposed to, *e. g.*, a perfect 3D plastic model of it. That is, no amount of congruence between representation and represented can ensure the kind of aboutness we demand of the contents of our mental states.

STRUCTURAL SIMILARITY is afflicted by one of the main problems that the right naturalistic account of content will have to solve: according to this theory, MMF represents my mother's face. After all. it does bear the right kind of structural similarity with it. But, unfortunately, MMF *also represents* every good enough 3D model, photograph, etc. of her face. After all, it bears the right kind of structural similarity with those things as well. So, contrary to what we wanted, the content of MMF is highly disjunctive: *my mother's face, or a 3D model thereof, or...* Instead, we need an account that allows MMF's content to be, solely, my mother's face. This *Disjunction Problem* must be solved.

DISJUNCTION PROBLEM: A theory of content suffers from the Disjunction Problem if the content attributions it warrants are highly disjunctive, even in cases where the common-sense content attribution is not.

A physical relation that seems more promising as a building block for content is *Being caused by*. Indeed, it looks as if part of what makes

2 It may be defended that the British Empiricists would have chosen this kind of properties.

a mental representation be about my mother is that it is her that causes it[3]. We might, then, try out the following simple causal account:

SIMPLE CAUSAL ACCOUNT - MMF: A mental state MMF is an Idea of my mother's face only if it is caused to token only by her face.

In SIMPLE CAUSAL ACCOUNT - MMF we are talking of the mental state *type* MMF; something that could, in principle, be tokened in different circumstances. As presented, SIMPLE CAUSAL ACCOUNT - MMF is afflicted by the Disjunction Problem, although it will be more informative to see the same issue under a different aspect: according to this simple causal account, states such as MMF cannot *misrepresent*. Suppose that a token of MMF is caused by a 3D model of my mother's face. We may want to say that the mental state in question has been fooled by the model's similarity to my mother's face, and that it has been *erroneously* tokened. But we cannot say that, because, in those circumstances, the token mental state in question ceases to be an Idea of my mother's face -as it has not been caused by it. This *Error Problem* is equally urgent.

ERROR PROBLEM: A theory of content suffers from the Error Problem if it never warrants a content attributions such that the contentful state in question is a misrepresentation.

The strategy deployed in SIMPLE CAUSAL ACCOUNT - MMF is letting the desired content fix the nature of the mental state -if a mental state is caused by something different, it is not ruled to be a contentful Idea. Another possibility -the one that, for expository purposes, is usually considered as the simplest version of causal-informational accounts of content- is taking the fact that a mental state is contentful to be independently fixed, and then letting the causes of its tokens fix its content. In general, an account of the content of mental states along these lines, such as

SIMPLE CAUSAL ACCOUNT - GENERAL: A mental state M is an Idea of whatever causes it to token.

cannot account for the possibility of error. Every mental state is about whatever causes it. If a mental state that, we might want to say, is an Idea of cat is caused to token by vaguely feline lumps, we are forced to say that it is also an Idea of vaguely feline lumps.

In the following subsection I will quickly survey some contemporary efforts to refine simple causal accounts, such that the resulting theory provides for the possibility of misrepresentation and does not attribute highly disjunctive contents in cases in which such attributions are counterintuitive.

1.1.2 *Somewhat More Complicated Causal Accounts*

3 We will need to qualify these remarks in due time. For example, my considered opinion is, rather, that what makes a mental perception be about my mother is that it is produced by a system that has to produce individual-involving representations, and that has produced my perception in the presence of a cue related to my mother in the right way.
For the time being, though, the simpler causal story is more informative, and better suited to my current introductory purposes.

STAMPE Stampe (1977) is customarily regarded as the first of contemporary efforts of analising intentionality in causal terms[4]. Stampe recovers from STRUCTURAL SIMILARITY the idea that an isomorphism between a set of properties of the representation and a corresponding set of properties of the represented is essential to one thing having the other as content. The twist in Stampe's account is that, for a genuine representation relation to exist, the latter must bear the right causal relation with the former:

> The causal relation we have in mind is one that holds between a set of properties $F(f_1...f_n)$ of the thing (O) represented, and a set of properties $\Phi(\phi_1...\phi_n)$ of the representation (R). Stampe (1977, p. 85)

Assume that R represents O in virtue of such a causal relation holding between them. This already takes care of much of the disjunctivity that afflicted content attributions according to STRUCTURAL SIMILARITY: suppose we wish to attribute a content to a photograph of my mother's face. There is, indeed, an isomorphism between certain properties of the photograph and certain properties of her face. If it is my mother's face's facial features that have caused the properties of the photograph that bear the relevant isomorphism with them, then the photograph represents my mother's face. As we have noted above, being isomorphic with my mother's face in the relevant way is, *eo ipso*, being isomorphic with a great many other things: 3D models, maybe, or my mother's twin sister's face. All of these things (which are part of what the photograph represents according to STRUCTURAL SIMILARITY) are ruled out by the causal condition: none of them have caused the photograph to have the properties it has.

This is already a partial solution to the Disjunction Problem, but there are many other sources of candidates for the role of represented object. As Stampe points out, if there is an isomorphism between the object represented and the photograph, there is surely going to be another isomorphism (a better one, actually) between the photograph and the impressed plate. But the photograph is only a photograph of my mother, not of the plate.

Now, as regards the Error Problem, this account so far is in the same situation as SIMPLE CAUSAL ACCOUNT - MMF. Consider, as a candidate for the role of representation, the stump a certain tree left after a lumberjack cut it; and the age of the tree at the moment of its death as a candidate for the content of that representation. There is, indeed, an isomorphism of the right kind between a certain property of the stump and the age at the time of its death: if the stump has n rings, the tree had n years, and, furthermore, it is the tree's having n years that has caused the stump's having n rings. Thus, the stump represents the tree's age at its death[5]. Now, suppose that the tree has sixty rings, but lived sixty-one years -one of them under a terrible draught. Stampe's theory -if left at this state- would not count the

4 This is, of course, little more than a stipulation useful for expository purposes. It is safe to say that Stampe's account would not have existed without e. g., Grice's notion of natural meaning (Grice (1957)) or the aforementioned groundbreaking lectures and papers by Kripke and Putnam.

5 Stampe in fact writes "that the stump shows sixty rings indicates that the tree was sixty years old" (*op. cit.*, p. 88). Here, apparently, it is states of affairs and not objects that represent. In the main text I'm sticking to objects (such as tree stumps) and not states of affairs as the right candidates for the role of representations. This seems closer to Stampe's official position.

stump as a representation in this case, given that it's not the age of the tree that has caused the number of rings. The system number-of-rings/age-of-the-tree is uncapable of showing misrepresentation. The Error Problem again.

Stampe makes an appeal to the function of the representing device in solving these two problems: that it is mother and not plate that the photograph is a photograph of depends on the function of the mechanism producing it (*op. cit.*, p. 83). Moreover, the possibility of misrepresentation is solved by adding a clause of fidelity conditions to the causal account, such that, *e. g.*, *if fidelity conditions hold*, years of tree life cause number of rings in the stump. This fidelity conditions are supposed to be the conditions under which "a functional system" functions well (*op. cit.*, p. 90)

We can, then, summarise Stampe's final proposal thus:

STAMPE: A representation R is about a thing represented O if and only if the right causal relation holds between a set of properties $F(f_1...f_n)$ of the latter and a set of properties $\Phi(\phi_1...\phi_n)$ of the former, if fidelity conditions hold.

This very insightful proposal suffers from some of the shortcomings typical of pioneering theories. First and foremost, it is little more than programmatic in several crucial respects. For example, it remains to be seen how the conditions of well-functioning of functional systems can provide for the possibility of misrepresentations; in particular, it is unlikely that the tree-age/stump system has any function that may help fix fidelity conditions for the reading of ages out of numbers of rings. Unlike photographs, trees are not *supposed* to keep track of their age in their rings.

It is also unclear how the teleological properties of photographs may -in a naturalistically acceptable manner- help track down one among the many isomorphic candidates for the object represented. It is not that this cannot be done; it's rather that Stampe's theory does not comment on it. It doesn't comment either on how should we ascertain, in STAMPE, what counts as "the right causal relation".

In a second level of importance, Stampe's theory is too heavily based on the model of photographs -hence the appeal to isomorphisms. It remains to be seen how STAMPE deals with, among other things, contentful entities such as the concept WATER -in which way is this concept to be isomorphic to water[6], and in what way such a isomorphic structure is to mediate our water-involving thinkings?

DRETSKE CIRCA 1981 In his (1981) Dretske offers a theory that makes content depend on information:

> Structure S has the fact that t is F as its *semantic content* = S carries the information that t is F in digital form. Dretske (1981, p. 177)

Where S carries the information that t is F in digital form iff it is the most specific piece of information that it carries about t (*ibid.*), and S carries the information that t is F iff $P(Ft|S) = 1$ (*ibid.*, p. 65).

The theory, as it stands, falls prey to the Disjunction Problem. Consider a mental state M the content of which we want to say that it is

6 In section 4.1 I discuss another, deep problem for theories that think of objects, as opposed to states of affairs, as the primary targets of representations.

There is a fly around. According to the passage just quoted, we may only attribute such a content if the probability of there being a fly around conditional on M being tokened is 1. So, if there is the slightest probability that M is tokened in the absence of a fly around -maybe in the presence of a black speck-, the desired content attribution will be unwarranted by the theory. Rather, the right semantic content attribution will have to be *There is a fly or a black speck around*.

The answer Drestke (*ibid*., p. 193f) suggests is postulating the existence of a learning period for each representation. It is during that learning period that the content gets fixed and, so, post-learning tokenings of the representation may well misrepresent. The final proposal may be rendered thus:

EARLY DRETSKE: Structure S has the fact that t is F as its *semantic content* = S carried the information that t is F in digital form during its learning period.

As Dretske came to see later on, this is not a very good way to account for misrepresentation. On the one hand, there is no principled way to distinguish learning and learned phases in the use of a representation. On the other hand, it is extremely implausible that there is any period at all in which any one representation perfectly covaries with its content. Rather, it seems that throughout its whole existence any representation will be erroneously tokened some times. It has to be possible to have a workable notion of misrepresentation even if this is the case[7].

We are going to spend some time (in section 1.2) discussing a theory in the spirit of Dretske's later theoretical efforts, and then other theories with a similar so-called *teleosemantic* outlook; but, before that, I wish to review quickly a more contemporary approach: Rupert's *causal-developmental theory*.

RUPERT In his (2008, p. 362f), Rupert defends the following theory of simple, atomic representations:

> If a subject S bears no content-fixing intentions toward R, and R is an atomic mental representation (i.e., not a compound of two or more other mental representations), then R has as its extension the members of natural kind Q if and only if members of Q are more efficient in their causing of R in S than are the members of any other natural kind.

Where the efficiency in question is to be calculated thus:

> For each natural kind or property Qi, calculate its PRF [past relative frequency] relative to R: divide the number of times an instantiation of Qi has caused S to token R by the number of times an instantiation of Qi has caused S to token any mental representation whatever. Then make an ordinal comparison of all Qj relative to that particular R; R's content is the Qj with the highest PRF relative to R. (*Ibid*.)

This is supposed to apply primarily to primitive representations (hence the antecedent of the conditional), the idea being that sophisticated content-crunching engines such as human minds can and do override facts about efficient causation with their intentional behaviour. So,

7 For a more sympathetic view of the learning-period strategy see Sterelny (1990).

for such primitive representations -*e. g.*, the mental states of frogs, or newborns- the property or natural kind that, in the long run and after the PRFs have reached their stable values, causes them to token is their content. This account is obviously able to accomodate misrepresentation -the content is the property with the highest PRF, which means that there are *other* properties with lower PRF that also cause the representation to token- and does not attribute highly disjunctive contents -these are plausibly ruled out by the constraint that contents be *natural* properties or kinds.

Rupert's causal-developmental theory is similar to what has been called *producer semantics* (see 1.3.2). One problem with these theories (the *Input Problem*, to be discussed in 1.3.2 as well) is that they warrant content attributions that are as close as possible to the firing patterns of the representation in question. In Rupert's version, whatever is most efficient in getting the representation to fire is its content. This, sometimes, provides counterintuitive content attributions. Take, for example, the case of *supernormal stimuli*:

> [A supernormal stimulus is an] artificial stimulus that produces in an animal a response that is stronger than would be evoked by the natural stimulus it resembles. For example, in some birds incubation behaviour is stimulated by the presence of an egg, and the larger the egg the stronger the stimulus; in such birds a very large artificial egg may be incubated in preference to a much smaller real egg. Allaby (2009)

Supernormal stimuli are more effective than, say, eggs in causing a certain subject to token a representation that, pretheoretically, we would like to identify as the perceptual concept EGG. In cases in which supernormal stimuli are possible, Rupert's theory will, counterintuitively, count the supernormal stimulus as the content of the representation in question. So, the brain state in the bird that causes it to initiate an incubation routine in the presence of a normal egg will not have the content, say, *There is an egg there* but, rather, *There is a very large artificial egg there*. Surely, this is implausible.

There is a way out of this unacceptable conclusion which does not look very promising: the causal-developmental theorist may retort that large artificial eggs do not form a natural kind, and thus are not candidates for being the content of a representation. While this may well be so[8], nothing prevents that some natural eggs (of, say, larger birds) have the features of supernormal stimuli. In such a case, the rejoinder is unsuccessful.

An apparently better answer is that these representations are there to help birds incubate eggs and reproduce; this is their biological function and that is why their content involves (healthy, fecundated) eggs and not supernormal stimuli, even if the latter are much better in triggering the representation than the former. Answers along these lines come from a familiy of theories commonly called *teleosemantic*, which make content depend on historical properties of representations -particularly those having to do with them, or their producers, having been selected through natural selection. I believe this is the most promising approach,

8 Although see Millikan (2000) and below in 1.4.5 for notions of natural-kindhood that would plausibly count some types of artificial eggs in.

and for the rest of the chapter I will concentrate on teleosemantic accounts of content[9].

Millikan's *biosemantics* is, of course, the best known and most sophisticated of the teleosemantic accounts available but, for starters, it will make sense to consider a simpler theory that recovers many of the Millikanian insights[10]: a somewhat updated version of Dretske's second take on the problem, *circa* 1986. We will see that a particularly insidious version of the Disjunction Problem can be used to cast doubt on teleosemantics in general; my own positive proposal is designed to solve this insidious *Indeterminacy Problem*.

1.2 A (BETTER) DRETSKEAN THEORY OF CONTENT

In this section, I introduce a concrete proposal about the content of some simple states that follows closely the one advanced in Dretske (1988). It is not *Dretske's* theory, mind you: I have tweaked several minor points to bring it closer to other contemporary accounts, and maximise its usefulness as a springboard for discussion. I will signal such tweakings as we go along.

A Representational System, according to Dretske is

> any system whose *function* it is to *indicate* how things stand with respect to some other object, condition or magnitude. Dretske (1988, p. 52, my emphasis)

As with the other accounts reviewed above, Dretske's definition should be understood as a step towards the naturalisation of the paradigmatically intentional notion of *representation*. The idea is to offer an analysis of this and cognate notions such as *content* into others which, it is hoped, may be tractably analysed in thoroughly non-intentional terms. It has been doubted, in this connection, that *indicate* and *function* can be given naturalistically unobjectionable analyses: maybe *indicating* is too close to *meaning* to be of any use, and scruples about the use of teleological idioms in science have, of course, a venerable history. In order to advance from Dretske's definition to a fully-naturalistic account of content, therefore, we need a non-intentional treatment of those two notions. I will discuss them now in turns.

One final introductory remark before that: I will be only concerning myself with extremely simple examples of contentful states. So, throughout this chapter I will assume that we are dealing with an innate mental mechanism, M, part of the cognitive setup of a subject S, which is always in one out of two possible states: *on* or *off*. My ambition is, simply, to offer a set of conditions for attributing the state-type [M's being *on*] with contents of the kind *There is an F around* in a way immune to the Indeterminacy Problem (to be shortly presented). How do the lessons we may learn from this kind of examples carry over to more sophisticated contentful states is matter for the rest of the chapters of this part I, specially chapter 4.

9 A brief discussion of the way in which the account I will favour accomodates the possibility of supernormal stimuli may be found in chapter 4, footnote 25.

10 Not by chance: Dretske's account is posterior and based on Millikan's first exposition of her theory in her (1984), as he fully recognises in his (1988).

1.2.1 *Indication*

We have seen in subsection 1.1.2 how Dretske suggests to analyse the notion of content in terms of the notion of information. The appeal to indication in the passage just quoted is doing a similar job. A bathroom scale indicates weight in virtue of the following fact: the position of the pointer of the scale *causally covaries* with the weight of the mass placed on the scale. The number of rings in a tree stump indicate how many years the tree has lived in virtue of the following fact: the number of rings in tree stumps *causally covary* with the number of revolutions of the Earth around the Sun during the lifetime of the tree. Indicators have a number of possible outputs (in our examples: pointer positions and numbers of rings) which causally covary with a number of possible states of the indicated system (e. g. different weights, different lifespans).

For mechanisms as simple as M, a straightforward way of understanding the indication relation is the following:

INDICATION: A mechanism M's going *on* indicates instantiations of a property F around S iff

I1: $P(F|M \text{ is } on) > P(F)$[11] and

I2: The difference in probabilities in I1 is causally grounded.

This is one of the points in which the Dretskean account I am introducing is not Dretske's account. For Dretske, there is indication only if $P(F|on) = 1$ (see above). Most theorists nowadays agree that this proposal, motivated by Dretske's views on epistemology, is unduly restrictive: in most cases positive correlation (understood as in I1) is enough[12].

Clause I1 means that the probability of F being instantiated around S, conditional on M being in its *on* position, is greater than the unconditional probability of F being instantiated around S[13][14]. These

11 Or, equivalently, $P(M \text{ is } on|F) > P(M \text{ is } on)$, if $P(M \text{ is } on) > 0$. Remember Bayes' theorem: $P(F)\,P(M \text{ is } on|F) = P(M \text{ is } on)\,P(F|M \text{ is } on)$. The formulation I've used in INDICATION is most congenial with Dretske's original idea (that an indicator causally covaries with the indicated). On the other hand, it lends itself less straight-forwardly to calculation of mean fitness values for a state (see below). In the sequel I'll be often using the condition $P(M \text{ is } on|F) > P(M \text{ is } on)$.

12 For discussion of this point cf. Shea (2007).

13 *Around S* is loose talk, of course, and could be made more precise in a number of different ways. Further refinement will be introduced in 4.2.

14 Millikan, in Millikan (2007), has raised doubts about the possibility of characterising the domain upon which to calculate probabilities such as those in I1 in a non-circular, yet useful way -the famous *reference-class problem*. I cannot go into a full discussion of this criticism, but I wish to suggest that some of her complaints can be assuaged by concentrating on the indication relations in which *the particular token of the mechanism* M *that S has* enters, instead of trying to characterise what Ms *in general* indicate.
We may help ourselves to a methodological fiction according to which, once the token of M in question has lived its whole life, and the actual frequency of coincidence of its being *on* with F being instantiated around S has been established, we rewind up to the point in time in which M started existing and press *play* again. We record the second actual frequency of coinstantiation and rewind again. After doing this a number of times, the mean frequency of coincidence will start converging to a value, which is $P(F|M \text{ is } on)$. We cannot actually *do* this, of course, which means that there may be an epistemic problem of how to know which is the right *assignment* of probabilities (cf. Gillies (2000, p. 813f); but, as far as I can see, these problems do not carry over to the metaphysical question of what it is that probabilities *are*. I am after an account of the metaphysics of content, not of the (very interesting, but different) problem of how to know that the account applies.
So if, in Dretske's famous example of the anaerobic, magneto-sensitive bacterium, we have two populations of bacteria, living in the northern and the southern hemisphere, do their

probabilities are to be understood as being objective, and not as tracking the opinions of some epistemic agent.

I2 is there to honour Dretske's request that the correlation between M's being *on* and instantiations of F be *causally grounded*. Some situations, among many others, in which an indicator-indicated relation complies with I2 are:

- The instantiation of F around S causes M to go *on*.
- The instantiation of a certain property E causes both an instantiation of F around S and M to go *on*[15].

A situation that does not so comply, on the other hand, is:

- Instantiations of F around S have happened to coincide, by a most extravagant coincidence, with M's being *on*.

The first building block on the way to the Indeterminacy Problem comes at this point. According to INDICATION, a state may, and most of the times will, indicate indefinitely many properties. For instance, a machine than bleeps if the person in front of it is more than 2 m high indicates the property of *Being more than 2 m high*, but also indicates *Having rather tall parents*. This is because the probability of the parents of the person in front of the machine being rather tall, conditional on it bleeping, is higher than the unconditional probability of the parents of the person in front of the machine being rather tall. For similar reasons, it indicates a great many other properties: *Making a decent basketball player*, *Using shoes above size 9*, etc.

1.2.2 *Function*

A difference between bathroom scales and tree stumps is that the former, but not the latter, may *misrepresent* something or other. My scale may give my weight wrongly, but it makes no sense to see to say that a tree stump gives the age of the tree wrongly[16]. We may be misled by the number of tree rings into thinking that the tree had a different age than it had when cut, but this is not to say that the stump is wrong or right.

A crucial insight exploited by Dretske -following similar ideas by Millikan- is that something may be a (mis)representation only if it has,

magnetosomes indicate the magnetic North and South respectively? According to INDICATION , this question must be posed independently for each individual magnetosome, throughout its life. The relative size of the northern and southern populations play no role whatsoever in ascertaining it.
In chapter 5 I will give reasons to think that this method of rolling back to a point in the past can help us know fully objective states of affairs regarding the counterfactual probabilities of M's behaviour being one way or another. All in all, this is congenial to *long-run propensity* theories of probability which, at least, are not obviously false -a useful review may be found in Gillies (2000, p. 126f). For a sophisticated view of objective single-case probabilities see Weiner and Belnap (2006) and references therein.

15 It is not clear whether the following third option would be accepted by Dretske as meeting I2: F is instantiated around S in some predictable sequence -maybe its being instantiated now causes the absence of instantiations in the following five minutes, which in turn causes another instantiation of F, etc. M also goes *on* and *off* in a similarly predictable pattern. These two patterns are such that F and M comply with I1.
As Millikan (2004) has pointed out, the indication relation going on in magnetobacteria is not causally-grounded in the strict sense: the Magnetic North indicates lower oxygen concentration in the Northern hemisphere not because there is any causal relation between these two gradients, but because both gradients stay put, so that one can usefully be used as a sign of the other. I intend INDICATION to be understood in a way such that this non-causal third kind of cases comply with I2.

16 *Contra* Stampe (1977), see above.

or has been produced by something that has, the function of indicating something or other. Now, bathroom scales have such a function because they have been designed by intentional agents such as us. But there are many central cases of representational states (our own perceptual states, for instance) which in all likelihood have not been designed at all. And, on the other hand, if in every unequivocal case of functional mechanism we had to appeal to the intentional states of a designer, functions would be unable to help us in the project of naturalising content. All in all, if our Dretskean proposal is to get off the ground, then, we need some account of how functions arise in Nature[17].

Before turning to the most popular naturalising account of functions, a *proviso* is in order. There is an important objection to the idea that a mechanism may have a function to indicate: the functions of a device are a subset of the *effects* of that device in a system, and indicating, instead, is passively standing in some conditional probability with regards to the instantiations of a property. If so, indicating cannot be the function of a state, something a state may be supposed to *do* -cf. Millikan (1993, p. 129), Neander (1991, p. 168), Kingsbury (2006). I think this objection is correct, as far as a theory of function goes. I also think that this issue is largely irrelevant for the project of naturalising semantics: *qua* content naturalisers, we are not, or need not be, interested in the correct account of biological functions; besides, the phenomenon described below, and codified in ETIOLOGICAL FUNCTION (namely, that a mechanism M's indicating instantiations of a property F may be part of an explanation of the selection for M) is clearly coherent, possible, and we have every reason to believe that it has also been actual a great many times. If the reader feels qualms about granting a function attribution to M upon such a history, she may substitute my uses of *function* in this context for *pseudo-* or *quasi-function*. Nothing hinges on this[18].

1.2.3 *Etiological Functions*

Hearts have, among others, the function of pumping blood and do not have the function of making thumping noises -even if the latter can be very useful for us, *e.g.*, as evidence in diagnosis. A very plausible proposal (mainly associated with Larry Wright, e.g., Wright (1973/1994), but also hinted at independently by a number of authors in the 70s, such as Ayala (1970)) is the so-called *etiological* approach[19]: the functions of such natural devices should be identified with certain explanations of the existence of those devices. So, Wright:

> [S]aying that the function of X is Z is saying at least that X is there *because* it does Z. Wright (1973/1994, p. 39)

17 Dretske does not commit himself to the account of functions I am about to introduce, or to any other.

18 Section 2.2 provides a fuller discussion of this point.

19 Other prominent approaches to the naturalisation of functions, about which I will have nothing to say, are the *dispositional* view -mainly associated with Cummins (1975)- and the *systemic* or *organizational* view -see Mossio et al. (forthcoming).
Schroeder (2004) has proposed to combine teleosemantics with a systemic approach to the naturalisation of functions, instead of the usual etiological approach. I cannot discuss his proposal here, but I do not find it very promising. According to Schroeder, the normativity of a representation depends on the fact that it belongs in a system with a goal state and a way to feedback information back into the system. I believe, for reasons that will be clear by the end of chapter 4, that it is highly unlikely that the goal-directedness of our cognitive system, disregarding its historical properties, can provide the right normativity for beliefs such as, say, "Bill Gates is tech savvy".

How does the fact that Xs do Z help explain the actual existence of an X in the relevant way -a way that bears on X's function[20]? There are a number of concrete proposals in the literature for unpacking this etiological insight, including some extremely sophisticated ones; among them Millikan (1984, , chapter 2f)[21], Millikan (2002), Neander (1995), Price (1998). To appreciate the force of the Indeterminacy Problem for semantics, nevertheless, it is best to keep the discussion at a conveniently uncommitted level. Here is a tenet that is common ground among most[22] etiological-function theorists:

SELECTION FROM FITNESS CONTRIBUTION: The existence of a biological trait or mechanism is explained in a way that bears on its function only if it has been selected[23] for.

This tenet helps ground a set of necessary and sufficient conditions for the attribution of etiological functions that makes the Indeterminacy Problem most conspicuous. A convenient way to provide such a set of conditions, for M to have the (pseudo-)function to indicate the instantiation of a property F, is using a simplified version of the apparatus of signal detection theory, as presented by Godfrey-Smith in Godfrey-Smith (1996):

So, how does M's being an indicator explains that it contributes to the fitness of its possessor, and gets thus selected? Assume that M indicates (remember: among many other properties) the instantiation of property F around S. The fact that M indicates F may help explain M's existence in the following way: in the simplest cases, M's going *on* causes, in its turn, S to initiate behaviour that is adequate for managing the presence of an F. Likewise, M's going *off* causes behaviour that is adequate for managing the absence of Fs[24]. This puts S in a better position to survive and reproduce than the one that M-lacking conspecifics enjoy. Eventually, this leads to the fixation of M in the population whence S comes.

We can summarise the effect that this adequate F-management has on the reproduction possibilities of S by assigning a *fitness value* to the four different states that result of the combination of M being *on* or *off*, and F being, or not, instantiated around it. Thus,

- M's being *on* whenever F is instantiated around the agent is a *hit*. We will assign this combination a fitness value w_{11} that, it

20 In Wright's quote X is a token. I'm using *X* now to talk about a type.

21 Millikan has repeatedly rejected that her theory of proper functions is influenced by, or in any way related to, Wright's. Although her theory is, in my view, a great improvement over his, I must say that I share the general opinion according to wish both are intimately related. This may be, of course, just another case of convergence of theories.

22 An important exception is Buller (1998).

23 In this dissertation I will assume that a classical approach to natural selection is correct; one according to which

> evolution by natural selection results whenever there is a population in which there is variation between individuals, which leads to those individuals having different numbers of offspring, and which is heritable to some extent. Godfrey-Smith (2009a, p. 4)

Furthermore, I will understand fitness simply as rate of reproduction. This is a substantive simplification, and I am sidestepping here many interesting philosophical discussions regarding the true nature of natural selection -see Fodor and Piattelli-Palmarini (2010) and Godfrey-Smith (2009a)- and of fitness. I believe my account to be more permissive of error in the classical view of selection than other similar accounts of content, but I will not discuss such issues here.

24 For simplicity, I'm helping myself to the assumption that doing nothing is a kind of behaviour.

is to be supposed, will be positive: the correct detection of an F will be coupled with some benefits (e. g., the successful hunting of prey, or escaping from predators) with an eventual impact in reproduction.

- M's being *off* whenever F is instantiated is a *miss* -fitness value w_{12}. In normal cases it will be detrimental: without recognising the presence of an F, the F-appropriate behaviour cannot kick in. We will lose a hunting opportunity, or may die devoured by an undetected predator.

- M's being *on* whenever F is not instantiated around the agent is a *false alarm* -fitness value w_{21}. It will also be negative in general. The relation between w_{12} and w_{21} will depend on the context in which F-detection takes place. If F is coupled with the presence of predators, normally $w_{21} > w_{12}$. That is, usually it will be more harmful for the fitness of the agent not to detect the presence of a predator than believing it is there when it is not. In the latter case only the resources spent in idle fleeing will be wasted; in the former, death is a likely outcome. *Vice versa* when F is coupled with, *e. g.*, the presence of prey, if prey is abundant. And,

- M's being *off* whenever F is not instantiated around the agent is a *correct rejection* -fitness value w_{22}. In normal cases w_{22} will be lower than w_{11} but higher than both w_{12} and w_{21}.

Thus we obtain the following *Fitness Matrix,* seen from the standpoint of the indication of F (FM_F):

$$FM_F = \begin{bmatrix} w^F_{11} & w^F_{12} \\ w^F_{21} & w^F_{22} \end{bmatrix}$$

On the other hand, we need to supplement the information that INDICATION provides about the relation between F and M, defining a complete *Indication Profile of F by* M (IP_F)[25]:

$$IP_F = \begin{bmatrix} P(on|F) & P(off|F) \\ P(on|\neg F) & P(off|\neg F) \end{bmatrix}$$

The Fitness Matrix, together with the Indication Profile, yields the following *Fitness Contribution* of M to its possessor, from the standpoint of the detection of F -that is, the weighted average of fitness values[26]:

$$\begin{aligned} FC_F &= P(F) \cdot \left[P(on|F)\, w^F_{11} + P(off|F)\, w^F_{12} \right] + \\ &+ P(\neg F) \cdot \left[P(on|\neg F)\, w^F_{21} + P(off|\neg F)\, w^F_{22} \right] \end{aligned}$$

where $P(F) = 1 - P(\neg F)$ is the objective unconditional probability of F being instantiated around the agent. We can use Fitness Contributions to offer the following set of conditions for the attribution to M of the etiological function of indicating F.

25 In the sequel I use $P(on|F)$ as an abbreviation of P (M is *on*$|F$).

26 I use the subindex *F* in FC_F to signify that we are dealing with M's fitness contribution as seen from the perspective of the indication of *F*.

ETIOLOGICAL FUNCTION: A token of the mechanism M in a subject S has the function of indicating the instantiation of F around S iff

- EF1: According to INDICATION, mechanisms of the type M have indicated the instantiation of F around its possessor in (a sufficient number of) S's recent ancestors.
- EF2: The Fitness Contribution, as seen from the perspective of indicating Fs, of the token of M in S's recent ancestors has been positive (that is, $FC_F > 0$), and this is part of an explanation of the fact that S has a token of M.
- EF3: (A sufficient number of) the conditional probabilities in IP_F are grounded.
- EF4: (A sufficient number of) the fitness values in FM_F are grounded.

I am making at least two assumptions here. First, that the history of the lineage of M is long enough, and pressures from the environment fine-grained enough, for M to end up having as fitness-contribution FC_F. In real cases M's fitness-contribution will approach FC_F, up to a certain level, and then move randomly in its proximity. Second, I are assuming that Fitness Matrix and Indication Profile are constant throughout the history of selection for M. Rather, it is likely that in more realistic scenarios such values change slightly as a result of changes in the environment, either random or directional. None of these two assumptions is crucial for my argument, although they do simplify the discussion.

Let me say something about the need for conditions EF1, EF2, EF3 and EF4 in turn.

EF1: This is necessary to comply with SELECTION FROM FITNESS CONTRIBUTION: if M's indicating F has contributed to the fitness of a sufficient number of S's ancestors, then it must have been the case that M has indicated F in those very ancestors. How many ancestors are a sufficient number of them is not susceptible of precise specification: clearly one ancestor is not enough and clearly all ancestors are enough; in between there is a large grey area. The reference to recent ancestors here and in the next condition is there to avoid once functional and now idle traits to count as functional -cf. Godfrey-Smith (1994), Griffiths (1993). Again, there probably are borderline cases of "recent enough".

EF2: This provides with the second necessary condition to comply with SELECTION FROM FITNESS CONTRIBUTION: indication must result in a positive Fitness Contribution, which must be part of the explanation of the actual existence of M in S.

EF3: It is important to see that how higher is $P(on|F)$ than $P(on)$ need not be the critical factor in selection for M. It is perfectly possible, *e. g.*, that a high rate of false alarms in M is compensated with a very low rate of misses, so that another mechanism M' such that $P(on'|F) > P(on|F) > P(on)$ ends up having a lower FC_F and is, therefore, not selected against M. This is why, in keeping with Dretske's requisites, if a low rate of false negatives has been instrumental for selection for M, then the probability of M's being OFF conditional on F's being instantiated or not must be causally grounded as well. How many of the four values in IP_F must be causally grounded depends on the actual

selection history of M: maybe all four are instrumental in selection, maybe less.

EF4: Again, in keeping with Dretske's requisites, the last bit of string for a tight causal package: the fact that, e. g., M's hits have a fitness value w_{11} must not be a bizarre coincidence. Imagine, for example, that the state M of an ant is very good at indicating pieces of metal and that, by most bizarre coincidence, on top of many pieces of metal there is a small lump of sugar. This case is such that hits have a high fitness value, but we wish to rule it out as a case in which the state M in the ant has the function of indicating metal[27]. This is the role of EF4. Again, how many of the four values in the Fitness Matrix must be grounded depends on the actual selection history of M.

Finally, it must be remembered that, with ETIOLOGICAL FUNCTION, we will only be able to warrant content attributions to states of mechanisms that have been selected for. Contentful states that come to exist during the life of a subject (*e. g.*, beliefs we form) need a more complicated treatment, and I deal with them in subsequent chapters.

1.2.4 *The Content of* M

Finally, the announced Dretskean proposal about the content of M's positives:

BETTER DRETSKE: M's being *on* has the content *There is an F around* if, according to ETIOLOGICAL FUNCTION, M has the function of indicating the instantiation of F around S.

BETTER DRETSKE, together with INDICATION and ETIOLOGICAL FUNCTION of course, completes the broadly Dretskean analysis of the content of M's being *on* promised at the beginning of this section.

1.3 THE INDETERMINACY PROBLEM

We are now in a position to introduce the Indeterminacy Problem. To use the standard (and, admittedly, rather tired) example in the philosophical literature, let me concentrate in a simplified frog, which I will call *Democritus*. We may tell the following idealised story about Democritus's mental mechanism M:

DEMOCRITUS AND THE CONTENT OF M'S BEING ON: Many generations ago, through random mutation, a mechanism M came into existence deep inside the brain of some ancient frog. Mechanisms of type M happen to indicate instantiations of the property of *Being a fly* around its possessor[28] .

They do this in the following way: an M fires whenever a black shadow of a determinate shape and size moves across the frog's retina. M also happens to provoke a response in the frog that

27 Its being a case in which M has the function of indicating *sugar* is already ruled out by EF1: the fact that $P(on|sugar) > P(on)$is not grounded, and this is one of the conditions for the indication relation to obtain, as I have defined it.
In section 2.1 I discuss whether we really wish to deem cases such as these as lacking function.

28 That is, *ex-hypothesi* each token of M has stood in the right kind of conditional probability vis-a-vis de property of *Being a fly*. See footnote 14.

makes it protract its tongue. Finally, M is hereditary. Thanks to these lucky causal powers, M tokens have helped frogs survive, by helping them hunt flies. This has meant that frogs with M have been fitter than frogs without, which in turn explains that M got fixated in that population. Nowadays, all frogs, and in particular Democritus, have M.

We may suppose that a more detailed account of the history of M allows us to calculate that $FC_{fly} > 0$, for many of Democritus's ancestors, in a way such that Democritus's token of M complies with ETIOLOGICAL FUNCTION. In that case, BETTER DRETSKE warrants a content attribution to M's[29] positives of *There is a fly around*.[30]

But, as Fodor (1990) points out (together with Neander (1995), Agar (1993), Rowlands (1997), and a great many others since) this is not the only option we have: there is an alternative, equally satisfactory explanation of the existence of M that leads in turn to a different attribution of etiological function and, finally, to a different content attribution:

DEMOCRITUS AND THE OTHER CONTENT OF M'S BEING ON: Many generations ago, a mechanism M came to be deep inside the brain of some ancient frog through random mutation. M happened to indicate instantiations of the property of *Being a black speck* around its possessor...

The rest of the explanation is the same as above. Indicating black specks contributes to the fitness of the possessor of M because, in the environment shared by Democritus and his ancestors, it has always been the case that a sufficient number of black specks were flies. And M does indicate black specks: the probability of M going *on* conditional on there being a black speck around S -which is causally grounded- is much higher than the unconditional probability of M going *on*.

Thus, ETIOLOGICAL FUNCTION also warrants the conclusion that M has the function of indicating black specks; finally, through BETTER DRETSKE , we are led to an attribution of content to M's positives of *There is a black speck around*. There is no fact of the matter as to which one of these two alternative explanations of the survival of M is the right one and, consequently, there is no principled reason to choose among the two alternative content attributions: hence the Indeterminacy Problem.

INDETERMINACY PROBLEM: A theory of content suffers from the Indeterminacy Problem if it warrants multiple content attributions to one mental state, even in cases where the common-sense content attribution is univocal.

1.3.1 *Different Descriptions, Same Fitness Contribution.*

I will now reconstruct the Indeterminacy Problem in a more formal fashion. It is maybe regrettable that nobody, as far as I am aware, has taken the time to do it before; all the more so as it provides crucial insights for solving the problem itself.

29 This is Democritus's token of M now.

30 Sometimes, informally, instead of using "the content of M's positives" or "the content of M's being *on*" or "the content of M's positives" I'll just talk of the content of M. Nevertheless, it should be reminded at all times that, rigourously, contents are attributed to types of states of mental mechanisms -M's being *on*, in our example.

Friends and foes of teleosemantics alike grant that M's indicating flies has been fitness conducive for Democritus and ancestors. Thus, if F is the property *Being a fly*, we have

$$FM_{fly} = \begin{bmatrix} w_{11}^F & w_{12}^F \\ w_{21}^F & w_{22}^F \end{bmatrix}$$

$$IP_{fly} = \begin{bmatrix} P(on|F) & P(off|F) \\ P(on|\neg F) & P(off|\neg F) \end{bmatrix}$$

$$\begin{aligned} FC_{fly} &= P(F) \cdot \left[P(on|F)\, w_{11}^F + P(off|F)\, w_{12}^F \right] + \\ &+ P(\neg F) \cdot \left[P(on|\neg F)\, w_{21}^F + P(off|\neg F)\, w_{22}^F \right] \end{aligned}$$

But, as I have shown above, there are many properties that M indicates. One of them is *Being a black speck*. In fact, M indicates black specks *better* than it indicates flies. That is, if G is *Being a black speck*,

$$IP_{black\,speck} = \begin{bmatrix} P(on|G) & P(off|G) \\ P(on|\neg G) & P(off|\neg G) \end{bmatrix}$$

where values in the diagonal of $IP_{black\,speck}$ are closer to 1 than those of IP_{fly}, and values outside the diagonal are closer to 0. On the other hand, this better indication profile is coupled with a different fitness matrix[31]:

- A correct positive of black specks, w_{11}^G, contributes less to fitness than a correct positive of flies, w_{11}^F : some of these positives will be the correct indication of a non-fly black speck. They will, therefore, have the fitness value of a false positive of flies, w_{21}^F. The percentage of such deleterious correct positives will depend on the percentage of black specks that are flies, thus:

$$w_{11}^G = \frac{1}{P(on|G)} \left[P(F|G)\, P(on|F)\, w_{11}^F + P(G \wedge \neg F|G)\, P(on|G \wedge \neg F)\, w_{21}^F \right]$$

  So, for example, w_{11}^G increases when the probability of M's firing in the presence of Gs decreases while the probability of M's firing in the presence of Fs is constant. That is, when the probability of *idle* firings in the presence of Gs goes down.

  w_{11}^G is a linear combination of w_{11}^F and w_{21}^F. Thus, if $w_{11}^F > w_{21}^F$ then $w_{11}^F > w_{11}^G$. A detailed calculation of w_{11}^G may be found in the appendix A.

- A false positive of black specks, w_{21}^G, is as deleterious as a false positive of flies, w_{21}^F: what is not a black speck, is definitely not a fly. $w_{21}^G = w_{21}^F$.

31 In the derivations below, I am assuming that $P(G|F) = 1$. That is, all flies are black specks. In fact, this assumption could be relaxed and we would still get the same result, if all the fitness-conduciveness of black specks comes from their being flies.

- A false negative of black specks, w^G_{12}, is less deleterious than a false negative of flies, w^F_{12}: some of these false negatives will be *correct* negatives of flies -if the black speck out there is not a fly. Again, the value of such beneficial false negatives depends on the percentage of black specks that are not flies, thus:

$$w^G_{12} = \frac{1}{P(off|G)}\left[P(F|G)\,P(off|F)\,w^F_{12} + P(G \wedge \neg F|G)\,P(off|G \wedge \neg F)\,w^F_{22}\right]$$

  And if $w^F_{22} > w^F_{12}$, $w^F_{12} < w^G_{12}$.

- Finally, a correct negative of black specks, w^G_{22} is as deleterious as a correct negative of flies, w^F_{22}.

The fitness contribution of M seen from the point of view of the indication of black specks is:

$$\begin{aligned} FC_{black\,speck} &= P(G)\cdot\left[P(on|G)\,w^G_{11} + P(off|G)\,w^G_{12}\right] + \\ &+ P(\neg G)\cdot\left[P(on|\neg G)\,w^G_{21} + P(off|\neg G)\,w^G_{22}\right] \end{aligned}$$

And finally, taking into account that

$$P(G \wedge \neg F)\,P(on|G \wedge \neg F) + P(\neg G)\,P(on|\neg G) = P(\neg F)\,P(on|\neg F)$$

and that

$$P(G)\,P(F|G) = P(F)$$

it can be shown[32], substituting the equivalences between fitness values calculated above, that

$$FC_{fly} = FC_{black\,speck}$$

This is, I submit, the root of the Indeterminacy Problem: regardless of how we choose to describe Democritus's situation -*i. e.*, from the perspective of M's indicating black specks or the perspective of M's indicating flies- we may obtain different pairs of Indication Profile and Fitness Matrix, but such that they converge in the same Fitness Contribution. And natural selection only cares about Fitness Contributions, not about from which point of view we look at it.

Assuming, as we must, that the rate of flies over black specks has remained approximately constant in Democritus's environment during selection for M, ETIOLOGICAL FUNCTION warrants an attribution of function to M of indicating black specks and BETTER DRETSKE warrants a content attribution of *There is black speck around*. It also warrants

32 The detailed derivation is in Appendix A.

the attribution *There is a fly around* and a great many others[33]. In Fodor's ingenious slogan,

> Darwin cares about how many flies you eat, but not what description you eat them under. Fodor (1990, p. 73)

1.3.2 *Problems In and Out, High and Low.*

In order to show the convergence of FCs that grounds the indeterminacy problem, I have put $w_{ij}^{black\,speck}$ in terms of w_{ij}^{fly}. But this should not mislead the reader into thinking that there is something intrinsically basic or fundamental about M's Fitness Matrix as seen from the perspective of indicating flies. Indeed, we very well may regard the fitness value associated with hits when indicating the property of *Being a fly* as a linear combination of the positive fitness contribution of hits when indicating *Being a fly with no frog-predator nearby* and the *negative* contribution of false alarms thereof (assuming that M also indicates this complicated property.) In any event, M's total fitness contribution will be the same, either as seen from the perspective of indicating black specks, flies or flies-far-from-predators.

Philosophers interested in the Indeterminacy Problem (*e. g.*, the above-cited Neander (1995), Price (1998) and Rowlands (1997)) have described two kinds of closely-related problems:

THE INPUT PROBLEM. First, what Price (op. cit) calls *The Input Problem*: we can formulate a *lowest* content attribution, finding the property L such that IP_L is as close to the 2x2 identity matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ as possible -I will sometimes call L the *lowest property*. Between the lowest attribution *There is an L around* and some (let us call it) "natural" attribution such as *There is a fly around* there lie indefinitely many attributions.

A teleosemanticist may defend that the right attribution for a state is the lowest, with, maybe, some additional constraints. This approach gives rise to what Rowlands (op. cit.) calls *stimulus-based* teleosemantic accounts, Papineau (op. cit.) calls *producer* teleosemantics, and Neander (op. cit., where she actually endorses it) calls *low church* teleosemantics. The Input Problem is the problem of stopping your theory from endorsing ever lower, to the point of totally implausible, content attributions.

For example, Neander's idea in her (1995, p. 129f) may be described thus:

NEANDER: A mechanism M's positives have the content *There is an F around* if, at the lowest level of functional analysis at which M is an unanalysed whole, it is supposed to detect Fs.

I will not get into the details of Neander's account, but the idea is that a mechanism such as Democritus's M occupies a certain role in the overall functioning of its possessor -it takes certain inputs, and serves certain outputs to other parts of Democritus. Moreover, M is part of

33 Incidentally, the fact that teleosemantic content attributions are multiply indetermined in the way we have shown invalidates Rowland's solution (cf. Rowlands (1997)) of distinguishing organismic and algorithmic proper functions attributions; the former is an attribution to Democritus, the latter to its mechanism M. For Rowland's proposal to get off the ground we should be able to identify indefinitely many *loci* to which attribute an etiological function, one for each competing attribution. This is not a promising project.

other, larger functional structures within Democritus and, in its turn, is composed of smaller parts. The level of description at which M is a standalone, unanalysed component of Democritus, is the level that fixes its content. In this case, the content is, then, whatever makes M go on; that is, black specks.

There are several problems with the account -and all other low church accounts. An important one is that it does not easily allow for the possibility of misrepresentation without malfunctioning -the possibility, that is, of tokening a wrong representation through no fault of one's own, only because the environment has not been collaborative.

But the problem that interests me most is that such low attributions are very counterintuitive. In the seminal (1959) paper that gave rise to the frog example, Lettvin reported feeling 'tempted' to call the convexity detectors in the *Rana Pipiens* eye 'bug perceivers'; not in vain such detectors work best

> when a dark object, smaller than a receptive field, enters that field, stops, and moves about intermittently thereafter. Lettvin (1959, p. 1951)

That is, Lettvin was distinguishing, as one should, between what causes a representation to token -small, dark objects in this case- and its content -bugs. We should strive at upholding this distinction.

THE OUTPUT PROBLEM. Second, the complementary *Output Problem*. We can formulate a *highest* content attribution, by looking for the property H such that FM_H has highest diagonal values and lowest anti-diagonal values -the *highest property*. Exactly what function should be maximised ($|w_{11} + w_{22}| - |w_{12} + w_{21}|$, or what) is unclear, but the intuition is that the highest content attribution involves the property that in fact accounts for the success of the possessor of M. It is not enough that M helps Democritus catch flies; they must also be nutritious flies, far-from-frog-predators flies, not-covered-with-frog-poison flies, etc. for the possessor of M to improve its fitness thanks to them. All of these increasingly complicated conditions for success amount to different proposals for content attribution[34].

The version of teleosemantics that defends the highest attribution as the right content-attribution (again, with some possible additional constraints.) gives rise to what Rowlands (op. cit.) calls *benefit-based* teleosemantic accounts, Papineau (op. cit.) calls *consumer* teleosemantics, and Neander calls *high church* teleosemantics. The most popular brands of teleosemantics are at least extensionally equivalent with some version of consumer teleosemantics as just characterised. In particular, in 2.2, once we are in a position to see why, I will show that Millikan's biosemantics is, indeed, subject to the Output Problem, at least in cases such as Democritus's.

What emerges from this discussion is a continuum of content attributions (see figure 1), from lowest to highest, depending on the choices we

34 It may be that $P(H|on) \leqslant P(H)$. That is, things be such that M does not indicate instantiations of the highest property; e. g., in cases in which a very fallible covariation is coupled with a very high fitness value of hits. In such a situation BETTER DRETSKE would not warrant the highest content attributions; reference to *the highest property* should be substituted with reference to *the highest property such that* M *indicates it*.
This is unlikely to be a problem: the fact that BETTER DRETSKE and related accounts warrant the highest content attribution is commonly perceived as a weakness, not a strong point, of the theory.

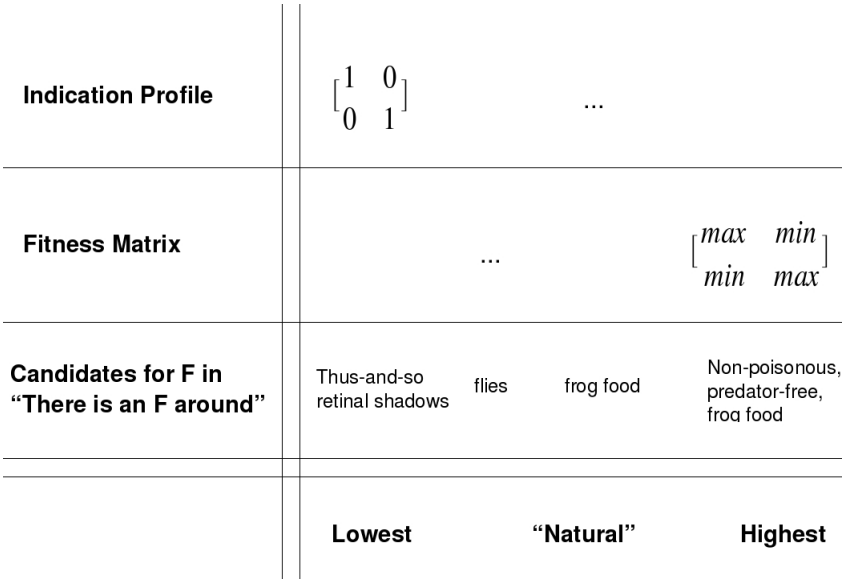

Figure 1: A Continuum of Content Attributions

make for Indication Profile or Fitness Matrix[35]. This was to be expected: the critical parameter behind selection is the Fitness Contribution of a state, which underdetermines the Indication Profile-Fitness Matrix pair that is needed to fix a content attribution. This extra degree of freedom gives rise to the multiple indeterminacy in content pointed out by foes of teleosemantics. Input and output versions of the Indeterminacy Problem are simply different perspectives on the same underlying issue[36]. More interestingly, we will see that the solution is equally homogeneous for both problems[37].

35 More rigorously: although, mathematically, there is a real continuum of Indication Profile - Fitness Matrix pairs, nothing ensures that, for each $\langle IP_i, FM_i \rangle$ pair, there will be a real-world property i such that M indicates i and M's Indication Profile from the perspective of indicating i is IP_i. The indetermination need not be as huge as the formalism suggests, although it will definitely be big enough for the Dretskean teleosemantic project to be in serious problems.

36 Other philosophers have described the same or closely related issues in different ways:

- The Distality Problem (*e. g.*, Price (1998), Ryder (2004), Stampe (1977)): How are we to decide whether the content of a mental state M involves a certain entity E, rather than entities causally downstream from E and upstream from M?
- The *qua* Problem (*e. g.*, Dretske (1988), Price (1998)): A content account suffers from the *qua* Problem in virtue of suffering from the Input Problem *or* the Output Problem.
- The Landslide Problem (Enç (2002)): Again, any of the Input or Output Problems. Enç's idea being that teleosemantic theories are on a landslide which takes them to ever lower -or ever higher- content attributions.

The sheer taxonomic variety present in the literature on the Indeterminacy Problem attests to the importance that philosophers accord to this problem. The Fitness Contribution, Indication Profile, Fitness Matrix framework can be used to describe all of these varieties easily. Admittedly, there is a kind of indeterminacy which slips through the framework: if there are two properties F and G such that both Fitness Matrix and Indication Profile seen from the perspective of indicating one are the same as seen from the perspective of indicating the other, the indeterminacy between the F-involving and the G-involving content cannot be capturedn in this framework. Fortunately, this seems like an unlikely case. Moreover, the account of content that I will be defending in this chapter takes care of this kind of indeterminacy just as it takes care of the rest.

37 Another point I'd like to stress, even if it's obvious, is that, given that the different indeterminacy problems can be given straightforward expression in the formalism, the formalism itself (in particular, the assumption that there is a Fitness Matrix coupled with each different Indication Profile) does not beg any questions against them.

1.4 THE CAUSAL BACK-OFFICE

The conclusion of the last section is, therefore, that there *is* an Indeterminacy Problem: traditional teleosemantics, indeed, warrant a multiply indetermined content attribution. Input and Output Problems are, both, consequences of the same background difficulty: if the fact that M covaries with some property F is evolutionarily useful, in most cases there will be an indefinitely high number of properties in the vicinity of F such that M covaries with them too, and this is evolutionarily useful for the same reasons.

The solution to the Indeterminacy Problem will emerge from spelling out what we mean by "the same reasons". We need, I submit, to pay attention to the causal grounds of the multiple indication relations that explain, each of them in an independently satisfactory manner, the evolutive success of M through its history. The general diagnosis will be that, although the causal underpinnings of attributions of functions to indicate are not enough to zero in on an univocal content attribution, among the facts that help explain the existence of a mental state M there are some, over and above those used in fixing function-attributions, that do pick out a unique content attribution. So far, this is trivial: there's a plethora of causal facts to choose from between the Big Bang and M. The difficult bit will be singling out of this plethora causal material that is, first, enough to zero in on a concrete F that is to figure in the content attribution *There is an F around* and, second, such that this singling out is plausibly regarded as grounding *content* attributions, not as merely providing an *ad hoc* solution to the Indeterminacy Problem.

I will presently elaborate on these sketchy remarks, first, by discussing in more detail which are the kinds of causal facts that must be in place for attributions of functions to indicate to be warranted. According to ETIOLOGICAL FUNCTION, we need two different kinds:

- (As per EF3) A causal explanation for (a sufficient number of) the conditional probabilities in IP_F.
- (As per EF4) A causal explanation for (a sufficient number of) the fitness values in FM_F.

I will discuss them in turn[38].

1.4.1 *Meeting EF3: The Causal Grounds of the Indication Profile*

Let us see which kinds of causal facts make EF3 true for a couple of functions-to-indicate attributions to Democritus's M. First, it will simplify the discussion to define M's *input*.

INPUT: M's input is the property F such that

1. M's indication profile of F is, as a matter of nomological necessity, the 2x2 identity matrix $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. That is, M is always *on* when F is instantiated around its possessor, and never otherwise.
2. FC_F is such that $w^F_{11} > w^F_{22}$.

38 EF1 and EF2 do not require the presence of causal grounds, just the very existence of the right Indication Profile and Fitness Matrix.

Clause 2 is necessary because, if there is one property F that complies with 1, there are two of them: we can redescribe M's performance relabelling *on* and *off* states the other way round. Under that description, $\neg F$ is M's input. Postulating that correct positives must be more fitness-contributing than correct negatives allow us to pick, out of M's two possible inputs, the most natural one: e. g., *Being a black speck* and not *Not being a black speck*. M's input individuates the lowest content attribution to M.

Now, let us see which kinds of causal facts meet EF3 for Democritus's M. I will do the exercise both for the function of indicating flies and the function of indicating black specks.

FROM THE PERSPECTIVE OF INDICATING FLIES. Let us suppose, for the sake of the example, that *Being a black speck* is M's input. Then, what causally grounds the elements in IP_{fly} is, on the one hand, the mechanism, partly constitutive of being M[39], that makes M fire as a matter of nomological necessity in the presence of a black speck. And, on the other hand, the fact that many flies are black specks.

Concretely, if $P(on|black\,speck) = 1$; $P(on|\neg black\,speck) = 0$, as we have just assumed, then

$$IP_{fly} = \begin{bmatrix} P(black\,speck|fly) & P(\neg black\,speck|fly) \\ P(black\,speck|\neg fly) & P(\neg black\,speck|\neg fly) \end{bmatrix}$$

So, whatever grounds the conditional probabilities around the agent of something being a fly given that it is a black speck helps ground also $P(on|fly)$. The facts that should be cited in this connection are the abundance of flies around Democritus, an estimation of the unconditional abundance of black specks, etc. In summary: in order to meet EF3 we need the external world to collaborate in providing a causal explanation of the fact that a sufficient number of black specks are flies, and a sufficient[40] number of flies are black specks. Maybe instantiations of *Being a fly* cause in some circumstances instantiations of *Being a black speck*, or vice versa, or the instantiation of both properties share a common cause, etc.

FROM THE PERSPECTIVE OF INDICATING BLACK SPECKS. Given that *Being a black speck* is M's input, we need no collaboration of the external world to make sure that $IP_{black\,speck}$ is the 2x2 identity matrix. The internal constitution of M is entirely to blame[41].

1.4.2 *Meeting EF4: The Causal Grounds of the Fitness Matrix*

FROM THE PERSPECTIVE OF INDICATING FLIES. M's output -those events caused by M's going *on*- involves Democritus's protracting his tongue and, subsequently, most of the times swallowing whatever it is

39 For a refinement and explanation of this claim of partial constitutiveness, see 2.5.1.

40 The 'sufficient' here can be given rigorous expression, with the help of the Fitness Contribution. The number of black specks that are flies will only be sufficient if it leads to a FC that explains the existence of M, as per EF2.

41 The caveat I raised in footnote 35 is in order here too: M's Indication Profile as seen from the perspective of property F will depend only of the mental state M itself only in case there exists a property such that it is M's input. This is a substantial metaphysical assumption that may not be always met -indeed, I am not sure that it is *ever* met. The point, though, is not very important for our current purposes, and we can help ourselves to the useful fiction that there is such as property.

that made M fire. Correct positives of flies are fitness-increasing because and when they lead to nutrient-intake, and such nutrient-intake has in turn caused differential reproduction in favour of each of Democritus's ancestors. Whatever it is that causally grounds, in each generation, the correlation between instantiations of the properties of *Being a fly* and *Being frog-food* (i.e., causally grounds $P(nutrient|fly)$) helps ground w^F_{11}. Remember: one causing the other, *vice versa*, both having common causes, etc. This explains that instances of *Being a fly* go together with fitness-increase; but we already had an explanation for the correlation between instances of *Being a fly* and M's going *on* (see above).

In case we need to provide causal grounds for the rest of values of FM_{fly} we will have to appeal to, *e. g.*, how heavily has the idle hunting of black specks been detrimental for the reproduction of Democritus's ancestors (w^F_{21}), and how severely the shortage of flies makes letting a fly pass detrimental for that reproduction (w^F_{12}).

FROM THE PERSPECTIVE OF INDICATING BLACK SPECKS. The same facts cited in grounding FM_{fly}, together with the causal grounding of the relation between *Being a fly* and *Being a black speck*, must be cited here. The causal story will be complicated by the fact that $FM_{black\,speck}$ values are linear combinations of FM_{fly}values, as calculated in 1.3.1.

1.4.3 *Etiological Function is Met. Now, What?*

Following the causal entangling of FM and IP for these two cases, we have been able to "decompose" the causal underpinnings of an attribution to M of the function to indicate Fs, warranted by ETIOLOGICAL FUNCTION. In summary, the world must provide with causal explanations for a number of conditional probabilities:

- $P(black\,speck|fly)$ and $P(\neg black\,speck|\neg fly)$,
- $P(nutrient|fly)$ and $P(\neg nutrient|\neg fly)$, etc.

Explaining why these properties tend to go together is enough to meet ETIOLOGICAL FUNCTION for each attribution of the function to indicate instantiations of the property F_i -that is, enough to explain the fact that M has the pair of Indication Profile and Fitness Matrix $\langle IP_i, FM_i\rangle$ for each property F_i. This is all it is sensible[42] to ask by way of causal underpinnings of function attributions.

Unfortunately, the Indeterminacy Problem has taught us that these causal grounds, the ones that underpin attributions of functions to indicate, are not enough to fix a unique content attribution for M...

1.4.4 *The Causal Grounds of the Stability of IP and FM Throughout Selection for* M

... Fortunately, on the other hand, barring bizarre thought-experiment scenarios, if there is selection for M a further causal ingredient will be in place, and this will be enough to zero in on a concrete natural kind to figure in the content-attribution. There will normally be an answer to the question: what makes M have such pairs of Indication Profile plus Fitness Matrix *across the generations needed for selection for* M? That is to say, there will normally be a causal explanation of the fact that the

42 Or, maybe, it goes beyond what is sensible. See below, section 2.1.

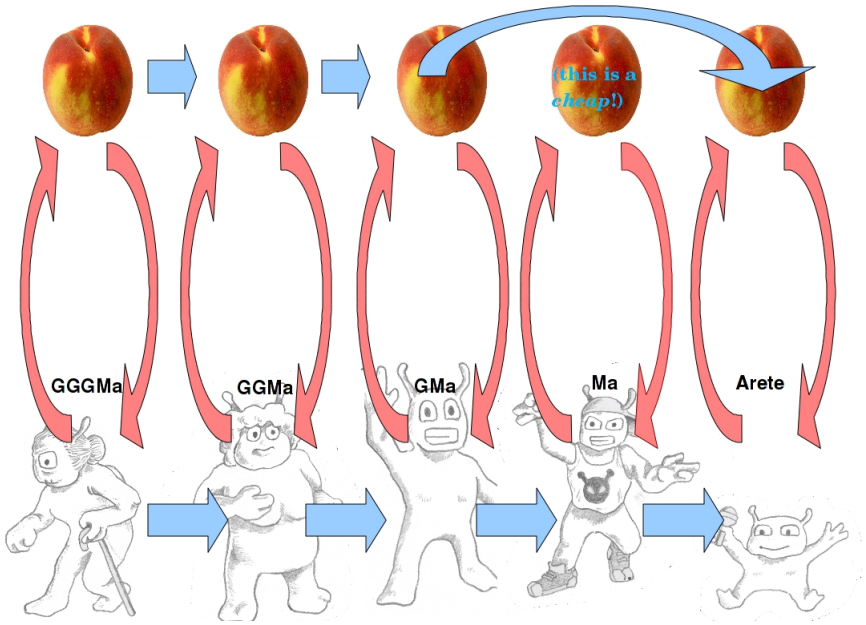

Figure 2: Arete and the Peach Tree

environment remains the same in the respects that ground M's Fitness Contribution.

To make vivid the relevance of this requirement for content attributions, let me introduce a situation in which it is *not* met.

ARETE AND THE PEACH TREE The mental state M is passed from mother on to daughter in Arete's family. M works like this: its input is a cluster of more-or-less peachy properties (peach shape, the particular shade of red-orange that peaches normally show, etc.), and it causes its possessor to grab and eat whatever she finds there where the input properties were instantiated. M has proven very useful for Arete's ancestors (Great Great Grandma, Great Grandma and Grandma), whose diet was chiefly based on peaches from a nearby peach tree.

But, most unexpectedly, one day the whole population where Grandma's daughter, Ma, belonged, was abducted in her sleep and deposited in Cheap Earth. Cheap Earth is just like Earth, except that no peaches are to be found there, but a very similar fruit, the *cheap*, genetically and causally unrelated to our peaches, although equally nutritious for creatures such as Arete. During her stay in Cheap Earth, Ma and the rest of M-endowed mutants (just a part of the population) survive better than many others just by doing what they have always done: grabbing and eating the peach-like object around. Some time later, the remnants of Ma's population, including her, is deposited back on Earth. Soon after that, she gives birth to Arete (*cf.* figure 2.)

The events depicted in ARETE AND THE PEACH TREE (we may suppose that these four generations stand in for many more) allow ETIOLOGICAL FUNCTION to warrant an attribution to M of a function to indicate: in *every* generation of Arete's family M's Fitness Contribution has been positive, and this explains that Arete has M. We may

discuss the composition of the causal grounds of $\langle IP_i, FM_i \rangle$ pairs in this example, just as we did in the previous subsection, and we would end up with the causal grounds of a number of conditional probabilities:

- P (nutrient|peach) and P (¬nutrient|¬peach),
- P (nutrient|cheap) and P (¬nutrient|¬cheap),
- P (peachy thing|peach) and P (¬peachy thing|¬peach),
- P (peachy thing|cheap) and P (¬peachy thing|¬cheap), etc.

The protest that some of these probabilities may not be adequately causally-grounded is misplaced: they all are. Cheaps and peaches, alternatively, ground these correlations for each generation. That is, we can attribute M with the function of indicating the property *Being a peachy thing*, according to ETIOLOGICAL FUNCTION: both EF3 and EF4 are met, for all generations.

Nevertheless, it is clear that there is something strange about Arete's story, more concretely around Ma's *akmé*. What is lacking there, and will be present in every actual case, is causal grounds for the *transgenerational* stability of the $\langle IP_i, FM_i \rangle$ pairs. A number of such pairs converge in a positive FC for each generation, and all of them are adequately causally-grounded. But we lack an explanation why such a thing is true, *e. g.*, *both* for Grandma *and* Ma.

1.4.5 *Homeostatic Property Clusters.*

So, a better explanation of the survival of M (and, I submit, the type of explanation that warrants content attribution) has to cite, apart from the collection of causal facts of the kind we have appealed to in subsections 1.4.1 and 1.4.2, whatever mechanisms account for the environment remaining such that M's Fitness Contribution is the same across generations.

Of course, bizarre cases such as Arete's mother never ever happen, and such mechanisms are in place whenever an M gets to be selected for its behaviour as an indicator. *E. g.*, genetic heritability among flies ground the constancy of $\langle IP_i, FM_i \rangle$ pairs in Democritus's lineage: the explanation that *Being a fly*, *Being a black speck* and *Being frog food* keep co-recurring throughout the time needed for selection for M to take place is that there is an *homeostatic mechanism* that keeps producing entities that present more or less those properties: the genetic reproduction of flies.

My proposal, finally, is that that we identify the natural kind F we talk about in content attributions to M such as *There is an F around* with the kind individuated by:

1. The relatively open cluster of properties the frequent co-instantiation of which grounds M's $\langle IP_i, FM_i \rangle$ pairs in each generation, together with
2. The *homeostatic mechanism* that explains that such cluster recurs in time, across the generations needed for selection for M.

Identifying certain natural kinds with a property cluster and its homeostatic mechanism, what is customarily called a *Homeostatic Property Cluster*, is not new, of course. The idea is originally suggested in Boyd

(1988), Boyd (1991), as a proposal for the kind of entities that normative properties such as *Being good* may be. Boyd's idea has had some success in contemporary philosophy of biology, as the right kind of entities for species (such as *fly*) to be; defences of this use of Homeostatic Property Clusters may be found in, *e. g.*, Brigandt (2009) and Wilson et al. (forthcoming). See also Richards (2008, p. 181) for a quick criticism of this project.

Neither is it new to use reasons why a cluster of properties recur as building blocks in naturalistic accounts of content. One prominent example is Millikan's development of the concept of *real kind* in her Millikan (2000) -with earlier versions of the same idea present since Millikan (1984). It is unclear, though, that a Millikanian real kind is individuated by something over and above the causal grounds that make some properties recur; it is unclear that the properties themselves play a role. Remarks such as

> [Real kinds] are things that *have* properties, rather than merely *being* properties. Millikan (2000, p. 15)

may or may not point to the idleness, in fixing real kinds, of the properties they normally have. Supposing that Millikan's account of real kinds is, more or less, the following:

REAL KIND: The causal grounds C that keep a set of properties recurring together accross time individuates a real kind.

There is at least the following worry with REAL KIND. We sometimes want to say that a species S_1 has evolved into another species S_2; most of the times environmental pressures will have driven the change but sometimes mere genetic drift is behind the speciation process. It is difficult to say, in this case, just what in the causal grounds that keep properties together is there to tell apart one species from the other. It looks, rather, as if *the same* causal processes are now responsible for keeping together a wholly disjunct set of properties, and that it is partly these new properties that take the two species apart. But, if so, proponents[43] of REAL KIND, which cannot help themselves to the disjoint sets of properties for kind-individuation purposes, have to count the two species as the same. This is an implausible result. Leaving speciation aside, in general, there may well be cases in which typical or frequent properties have some role in fixing the relevant real kind, even if not a criterial one.

Another similar idea put to use in a content naturalisation strategy is Ryder (2004)'s *sources of correlation*. Also Ryder seems to doubt whether taking the properties whose recurrence is causally grounded as part of what a real kind is. When he introduces the notion of 'unified property cluster' (2004, p. 213), sources of correlation individuate, together with 'a set of correlated properties', these clusters. In the subsequent discussion, though, he talks of correlation grounds alone as that which representations are about. It seems, thus, that his account is subject to the worry about speciation just presented.

In any event, what is new, I venture to say, is the result I have been arguing for: that something similar to Millikanian real kinds, which I will call Homeostatic Property Clusters (also HPCs from now on), solve the Indeterminacy Problem for naturalistic accounts of content. So that

43 Which, let me say again, may or may not include Millikan.

we do not need to rely on our intuitions about these entities, let me try to lay down as explicitly as possible what I will be meaning by HPC throughout this work[44].

HPC: Given a class of properties *P*, and a class of their instantiations, *SEED* (I will sometimes call this class a *seed* of the HPC), if there is a specialised homeostatic mechanism *SHM* that explains, in a certain domain *d*, a sufficient number of the instantiations in SEED of properties in P, then there is also the smallest class of properties *P'* such that SHM explains the fact that, in *d*, the properties in P' are frequently coinstantiated.

We define an HPC as the entity individuated by

- The set PI' of those among the instantiations of properties in P' that are explained by SHM (I will sometimes call this set the *property cluster* of the HPC), and
- SHM.

For simplicity we may understand *domain* as *spatiotemporal domain*; but we should keep in mind that domains are also constrained in other dimensions: ranges of temperatures, pressures, ion concentrations, etc. The domain *d* doesn't have to be particularly smooth; it may well be discontinuous and take any shape whatever, but we shouldn't think of it as including only the place and time (temperature, etc.) at which the instantiations in the Cluster occur. It is difficult to be more precise in this respect, but there seems to be a clear sense in which the domain of, *e. g.*, flyhood is not exhausted by the places in which actual flies have been; flies could have occupied a number of nearby places, and flown following other trajectories. Those are also part of the HPC's domain.

Let me explain the appeal to 'specialised homeostatic mechanisms'. A homeostatic mechanism is a type of causal processes that explains that the environment remains the same across some dimension in some respect. For example, in the case at hand, it explains that the density of instantiations of properties in P across space and time remains more or less constant, at least within the bounds of *d*. We need *specialised* homeostatic mechanisms because there are non-specific homeostatic mechanisms which explain the frequent coinstantiation of properties in *P*, but also of properties such as *Being an oxygen-rich portion of atmosphere* in *d*. For example, one such mechanism is the sum of every homeostatic mechanism that there will ever be. Such too general mechanisms will distort the HPC beyond recognition[45]. Another example; suppose the Gaia theory is correct, and the Earth is a giant ecosystem. There is a homeostatic mechanism that keeps it stable across time, and maybe the mechanism that keeps the density of flies constant across time is a submechanism of that one[46] So we need a way to distinguish relevant from irrelevant here. One way to do so is to choose the set of homeostatic mechanisms that explains the frequent coinstantiation of *fewest* properties -provided that it explains the coinstantiation of those in *P*. This is *SHM*.

44 What follows is, although obviously indebted to Millikan and Boyd, very much my own development of the notion of HPC. It is definitely not to be taken as a statement of Boyd's position, or of any other philosopher's.

45 This problem, by the way, although seldom if ever discussed, afflicts every theorist, from Stampe to Ryder, that makes covert or overt use of real kinds.

46 There are many interesting questions here about mechanism individuation which must remain outside the scope of this work. I hope to tackle them elsewhere.

Notice that there is no guarantee that the specialised homeostatic mechanism in question only explains the coinstantiation of properties in *P*; *P* may be a proper subset of that smallest set of properties. We want the smallest set *P'*, and not *P*, to individuate HPCs at least for the following reason. Imagine a situation in which a certain animal -a cat, say- is a prey to some -dogs- and a predator to others -mice. Suppose we construct the class of properties that enabled selection for a cat indicator in dogs (M_{dog}) and mice (M_{mouse}). It is likely that M_{dog} relies on a class that includes properties such as *Being nutritious for dogs*, while M_{mouse} relies on a class that includes properties such as *Being dangerous for mice*. We do not want to say that the HPC that mice detect is different from the HPC that dogs detect. This is effectively avoided by allowing instantiations of *P'*, and not of *P,* to individuate HPCs[47].

Boyd's theory of HPCs is customarily taken to be a somewhat deflationary account of real kindhood compared to the *traditional essence account* of real kinds. This other account has it that real kinds have a hidden essence -say, *Being* H_2O for water, or *Being the element with atomic number 79* for gold- such that all and only that which has that essence is part of the real kind in question. I wish the notion of homeostatic mechanisms to be understood in a way compatible with the essentiality of essences, when they exist. So, for example, a certain specialised homeostatic mechanism may have, as a proper part, the fact that certain molecular structures give rise in certain conditions to certain macroscopic properties. If these macroscopic properties are part of the cluster of an HPC, the fact that it is the molecular structure in question, say, M, which helps explain that properties recur is an essential property of the HPC. The twin HPC that is kept in place by molecule M' is a different one[48].

A HIERARCHY OF HPCS When, in the next chapers, we set about to explain the possibility of not selected, yet contentful, mental states, we will need *higher order HPCs*. Their definition is recursive.

HIGHER ORDER HPC:
1. An HPC is of order 1 iff none of the instantiations in the Cluster is of a property *Being an F*, where F is an HPC.
2. An HPC is of order $n+1$ iff any of the properties instantiations in the Cluster is of *Being an F*, where F is an n^{th} order HPC.

A higher order HPC provides a causal explanation of the fact that two or more (lower order) HPCs are frequently coinstantiated. For example, *car* is a high order HPC which explains the frequent coinstantiation of wheels and tyres -which, I am assuming here, are also HPCs. In 2.5 I discuss the difference between the kind of properties one may find in the cluster of 1^{st} order HPCs, and properties such as *Being an F*, where F is an HPC.

47 Thanks to Sònia Roca for discussion here.
48 There is further discussion of the relation between tradicional-essence real kinds and HPC in 2.5.2.

1.4.6 *The Content-Attribution Recipe*

In sum, I submit, in order to obtain the correct content attribution for simple mental states such as M's positives, we should follow the following recipe:

- First, we need the properties the correlation among which ground Indication Profile and Fitness Matrix pairs for M. These properties will be part of the cluster of the HPC.
- Second, we need the causal mechanism that explains the recurrence of said properties across generations of the possessors of M and, therefore, explains that M keeps being fitness-contributing across generations. This is the specialised homeostatic mechanism appealed to in HPC (e. g. fly-reproduction, plus whatever enabling environmental conditions are needed, such as the presence of oxygen, a fertile soil, etc.)

Once we have identified these different ingredients of the explanation of the existence of M, we may now give a content-attribution recipe:

THERE IS AN F AROUND: M's being *on* has the content *There is an F around* if[49]

TFA1: ETIOLOGICAL FUNCTION warrants, for a number $i \geqslant 1$ of properties G_i[50], the attribution to M of the etiological function of indicating the instantiation of G_i around its possessor S.

TFA2: The different pairs, $\langle IP_{G_i}, FM_{G_i} \rangle$, of Indication Profile and Fitness Matrix for each property G_i are grounded on the frequent co-instantiation of several properties in S's environment. (cf. subsections 1.4.1 and 1.4.2)

TFA3: The fact that $\langle IP_{G_i}, FM_{G_i} \rangle$ pairs remain the same across S's lineage is causally grounded on a specialised homeostatic mechanism *SHM* that explains[51] the recurrence of the properties appealed to in TFA2 in a certain domain *d* around M.

TFA4: F is the natural kind individuated by *SHM* and the smallest set of properties *P'* such that *SHM* explains the fact that, in *d*, the properties in *P'* are frequently coinstantiated.

The notion of *causally-grounded* in TFA3 is also technical:

CAUSALLY GROUNDED: A specialised homeostatic mechanism SHM causally grounds the persistence of $\langle IP_{G_i}, FM_{G_i} \rangle$ pairs of an indicator M across time only if:

49 Not "... and only if". We are just providing sufficient conditions for a state to have the content *There is an F around*, not necessary ones. None of our conscious mental states meets THERE IS AN F AROUND. In subsequent chapters I will enlarge the class of mental states covered by the theory -including states that have not been selected for. But, in general, I will only defend the existence of interesting -relevant, frequently instantiated, etc.- sets of sufficient conditions for contents to exist. I don't think much of the project of providing a set of necessary and sufficient conditions. In this, I share the qualified scepticism about the project of naturalising content vented by Godfrey-Smith (2004b), Godfrey-Smith (2004a).

50 These properties need not be properties of Fs.

51 That a particular SHM explains or not the recurrence of a cluster of properties may be a vague matter, with paradigmatic and borderline cases. There may be cases, that is, in which it is unclear whether a certain HPC is part of the content of a mental state, or whether the mental state is contentful at all. See, in 2.5.1, the discussion of disjunctive contents.

CG1: The domain in which SHM applies overlaps significantly with the domain in which M exists; also, a sizeable number of the property instantiations that have caused M to go *on* historically must be part of SEED.

CG2: SHM does not explain instantiations of the property of *Being* M, unless this means that SHM does not explain the instantiation of *any* property.

CG1 is there for reasons such as the following: if a number of physical replicas of flies, but causally unrelated to flies, appear near Democritus and start reproducing among themselves, although not with normal flies, and Democritus never feeds on them, we wish to avoid the content of M's being *on* to be *There's a fly-or-swampfly around*, where *swampflies* are these physical replicas. In general, the appeal to concrete property instantiations in the definition of HPC is there to avoid that HPC extensions include causally-unrelated swampreplicas. In CG1 we are fixing *which* HPC should be invoked in the content attributed.

CG2 is there to avoid M itself to be part of the Cluster. On the other hand, we need to admit cases in which a thought is directed to own's own thoughts -*e. g.*, a cogito thought, but not only those. So, if ruling out that the Cluster includes M empties the Cluster, we repeal the rule.

THERE IS AN F AROUND is, I submit, an account of the content of simple mental states that recovers all the advantages of proposals such as BETTER DRETSKE, without, crucially, falling prey to the Indeterminacy Problem.

To put the main idea under a different light, the proposal is that the content of these simple, innate mental states should not involve (as Dretskean teleosemantics wanted) useful properties such that there is an explanation why they can be detected but, rather, it should involve *the very structure* individuated by those properties and this explanation. This structure (the HPC) is not any random mereological sum of causal facts and property instantiations, but may be plausibly regarded as a real kind.

Before finishing this first chapter I will show that this is in fact a solution to the Indeterminacy Problem, and then I will take up the question why should we think of THERE IS AN F AROUND as offering *content* attributions.

1.5 THE INDETERMINACY PROBLEM SOLVED

A quick recap: one may provide indefinitely many properties such that, on the one hand, M indicates them and, on the other, this explains M's Fitness Contribution. For example, for Democritus: *flyish things*; *flies*; *black specks*, etc. Some of them M indicates very accurately (*i. e.*, they have an Indication Profile that is nearly the identity matrix). Some of them less so, but, according to BETTER DRETSKE, content attributions involving each of them are, all, warranted. Another bunch of candidates for warranted content-attributions surface if we focus on the properties that ground the fitness gains -that is, properties that maximise values in the diagonal of the Fitness Matrix: *nutritious stuff*, *frog food*, *unpoisoned frog food*, etc. Again, each of these properties ground a different $\langle \mathrm{IP}, \mathrm{FM} \rangle$ pair, but all of them amount to the same *FC*.

Now, the way to solve the Indeterminacy Problem: we have to look for the properties that explain that conditions EF3 and EF4 of ETIOLOG-

ICAL FUNCTION are met for these alternative content attributions. This class of properties will be the seed of the HPC -let us call it P.

If Democritus's M is to be contentful, there has to be a specialised homeostatic mechanism that has made properties in S recur around Democritus and its ancestors. In the normal case, this mechanism is fly reproduction, together with enabling environmental conditions -call it *FR*. FR is able to explain the recurrence of the properties in S in a certain spatial region *a* (that is, the region where flies are naturally present), and during a certain time *t* (from the time in which flies appeared to the future time in which they evolve to something else).

Now, according to THERE IS AN F AROUND, the HPC that must figure in the content attribution is the one individuated by *FR*, together with the set P' of all properties such that FR explains their recurrence in *a* and *t* (P will normally be a proper subset of P'):

> P' = {*flyish things*, *flies*, *black specks*, *frog food*, *unpoisoned frog food*...}

We happen to have a name in English for such an HPC: *fly*[52]. So the content of [M's being *on*] is *There is a fly around.*

As we have seen, the fact that M has indefinitely many functions to indicate -the root of the Indeterminacy Problem- is not a difficulty for THERE IS AN F AROUND: the HPC that must figure in the content attribution remains the same regardless of which such function-to-indicate we choose.

1.6 HOW IS THIS CONTENT?

Our inquiry on the naturalistic basis of content attributions began with accounts that propose a causal *analysans* which appears to have an intuitive claim to being a candidate for the content of representations. Thus, SIMPLE CAUSAL ACCOUNT - GENERAL: the things that cause representations do have a *prima facie* appeal as possible candidates for the role of contents, even if, on *secunda facie* reflection, the Error and Disjunction Problems have convinced us otherwise. Similarly, BETTER DRETSKE has intuitive appeal too: contents do seem the kinds of things that representations may have the function to indicate. Unfortunately, learning to live with the Indeterminacy Problem is simply not an option.

On the other hand, THERE IS AN F AROUND has none of these imposing problems. I have, I hope, provided sufficient reasons to think that it is not subject to the Indeterminacy Problem. The Error Problem is clearly not a difficulty for the account either: *e. g.*, any old black speck can cause Democritus's M to go *on* in the absence of a fly around. Neither is the Disjunction Problem: the content attributions warranted by the account are of the well-behaved, non-disjunctive form *There is an F around*, where F is an HPC[53]. The worry I wish to consider now is whether this resilience to problems against causal accounts of content is achieved at the cost of not being an account of *content*, at all. Indeed, the

52 Although this is not important for our present concerns, we may notice that *fly* is a higher (at least 2nd) order HPC: one of the properties in the cluster is *Being an individual fly*, which is another HPC. In this respect, *fly* is similar to *chair* and different from, say, *gold* or *water*. This additional structure in the former but not the latter HPCs is signalled in language by the use of count (as opposed to mass) common nouns. We will witness the additional structure doing some work in 4.2.

53 In section 2.5.1 I will provide conditions for the attribution of some disjunctive contents. Nevertheless, the emergence of systematic abilities to form and entertain disjunctive contents will have to wait until compositionality makes its appearance, in chapter 4.

involved condition for being the content of a representation according to THERE IS AN F AROUND does not seem to strike an intuitive chord the way, e. g., the appeal to functions to indicate does. I wish to argue that, after *secunda facie* reflection, it does.

The first thing to note is that THERE IS AN F AROUND, just like BETTER DRETSKE and SIMPLE CAUSAL ACCOUNT - GENERAL abide by the following general principle:

COMPRESSED EXPLANATION To provide a content attribution for a representation type R is to provide a compressed explanation of the existence of R in a sufficient number of cases.

COMPRESSED EXPLANATION may be thought of as a partial elucidation of the essence of content: perhaps, after sufficient reflection it is *a priori* true that content attributions are, in fact, in the business of explaining the existence of representations. At any rate, all of the theories we have considered in this chapter comply with it. Take SIMPLE CAUSAL ACCOUNT - GENERAL. Indeed, that which causes a representation to token (this is the content candidate according to the theory) figures in an explanation -the most proximal- of the existence of the caused representation. Or BETTER DRETSKE: at least under the etiological understanding of what functions are, to say that something is whatever a representation has the function to indicate (*i. e.*, its content) is to say that it figures in an important way in an explanation of the existence of the representation: had not the ancestors of the representation indicated it, the actual representation would not exist[54].

THERE IS AN F AROUND also abides by the principle. In fact, according to my theory, one may recover a substantial part of the story that explains the existence of a (simple, innate) representation just from its content attribution: the HPC that figures in a content attribution, indeed, explains how the emergence of Dretskean functions-to-indicate, that in turn explain the actual existence of the representation, was possible in the case at hand.

A second thing to note is that what explains the appeal of SIMPLE CAUSAL ACCOUNT - GENERAL is also present here: there is a real kind (an HPC) lying about in the environment, and it is this real kind that brings the representation into existence. At this level of description both theories are on a par. It is only that, we have found out, not any old way of causing the representation is sufficient to fix a content; rather, our investigation has led us to conclude that only a very concrete kind of intervention of the HPC in the events leading to the existence of the representation will secure the former a place in the content of the latter[55]. The theory I am proposing is the wised-up version of a simple causal account. To stress the crucial role that, if I am right, causal

54 As I have explained above, Dretske himself is not committed to the etiological theory of functions, or any other, being the right one. On the other hand, it is maybe telling that most contemporary accounts that rely on broadly teleosemantic insights also uphold etiological functions. A prominent counterexample to this is Cummins (1991). But Cummins himself points out that his account is not after the content of thoughts but rather after "the content of the representations of a computational system" (*id.*, p. 88).
So, even if Cummins's theory were correct, COMPRESSED EXPLANATION may still be true for the content of thoughts.

55 This is so for selected-for, simple representations. In subsequent chapters we will see that, in more sophisticated content-crunching creatures, the content of a representation is sometimes fixed simply by what first caused it to token -or something close to it. This is possible because the 'concrete kind of intervention of an HPC' is done at a different level, *e. g.*, in the selection for the mechanism that *creates* mechanisms such as Democritus's M.

explanations of the existence of representations have in fixing their content, I will call the theory to be developed in this work, of which THERE IS AN F AROUND is the first building block, *Etiosemantics*. I will establish in the next chapter -although it's probably apparent already- that etiosemantics is not *teleo*semantics: indeed, the causal explanation needed to fix content has as a *proper part* the causal explanation needed to fix the biological function of representations.

A final consideration in favour of etiosemantics comes from the kinds of attributions the theory makes. The account uses as content attributions the entities that keep together a large number of properties that are of use to the possessor of the representation -immediately useful ones such as *Being food*, and useful for the detection of the former, others such as *Being a black speck*. It is likely that the entity distinguished with this role will not lie at any of the two extremes that give rise to the Input and Output Problems. In Democritus's case, for instance, the entity is *fly*; one of the "natural" attributions we would have pretheoretically favoured.

All in all, I hope I have made a case for drawing the following tentative conclusion: THERE IS AN F AROUND, upon reflection, provides a candidate for the role of content that is as good from the intuitive perspective as other prominent causal theories of content. This first impression, I hope, will be confirmed once the rest of the Etiosemantics theory, to be developed in the following chapters of this part I, is in place.

On the other hand, THERE IS AN F AROUND does not suffer from the problems of Error, Disjunction or Indeterminacy. It is, thus (and tentatively) to be preferred to the other accounts discussed in this chapter. In the next chapter I chart more carefully the differences between etiosemantics and teleosemantics, paying close attention to Millikan's version. I will also be more explicit about some of the assumptions of the theory. Finally, I provide the etiosemantic answer to some well-know objections to teleosemantics.

2 ETIOSEMANTICS AND TELEOSEMANTICS

According to etiosemantics, it is the existence of HPCs enabling the emergence of functions-to-indicate that guarantees univocity of content. That there is an HPC around to play the role the theory demands, however, is not guaranteed by the fact that a mental mechanism M has an etiological function of the kind teleosemantics builds content-attributions on. In this respect, etiosemantics parts company with strict teleosemantics. Section 2.1 explores the kinds of situations in which teleosemantics and etiosemantics yield different predictions. I defend etiosemantic predictions by showing how two contemporary broadly teleosemantic theories fall prey to the Output Problem -Millikan's biosemantics- and the Indeterminacy Problem -Shea's infotel semantics as developed in Shea (2007). I also discuss Papineau's version of teleosemantics (2.4), and defend that it only solves the Indeterminacy Problem under a highly idiosyncratic understanding of what desires are.

After that I show how even the small fragment of etiosemantics I have introduced so far can accomodate several semantic phenomena (reference change, disjunctive contents, etc.), I review some metaphysical assumptions of the theory, and provide an answer to some well-known objections to teleosemantics.

2.1 CONTENTLESS INDICATORS

I have claimed that my proposal of content attribution for simple mental states parts company with teleosemantics, at least of the BETTER DRETSKE variety. One way to see this is with a situation in which ETIOLOGICAL FUNCTION warrants an attribution to M of the function to indicate instantiations of a property F around M's possessor, but THERE IS AN F AROUND does not warrant any content attribution to M's being *on*. It will have to be a thought experiment, though. Cases in which the states of a mechanism with the function to indicate are not contentful most probably have never existed and will never exist.

An easy way to provide an example of a contentless indicator is by modifying the ARETE AND THE PEACH TREE example, making Ma's case the norm rather than the exception. That is, ensuring that no HPC plays a role in the explanation of the existence of M:

MANY EARTHS, MANY PEACHES: Not just Ma's generation but, at time intervals that correspond roughly with *every* generation, the whole population in which she belongs has been, by chance, abducted and transported to a different planet. Great Great Grandma and her conspecifics to Earth_3, Great Grandma and hers to Earth_2, etc. Such planets are very similar to the Earth, except for their not having peaches but, rather, peaches_n -different kinds of fruit, similar in all respects to peaches, but such that for every i and j, peaches_i and peaches_j are genetically, and otherwise causally, unrelated.

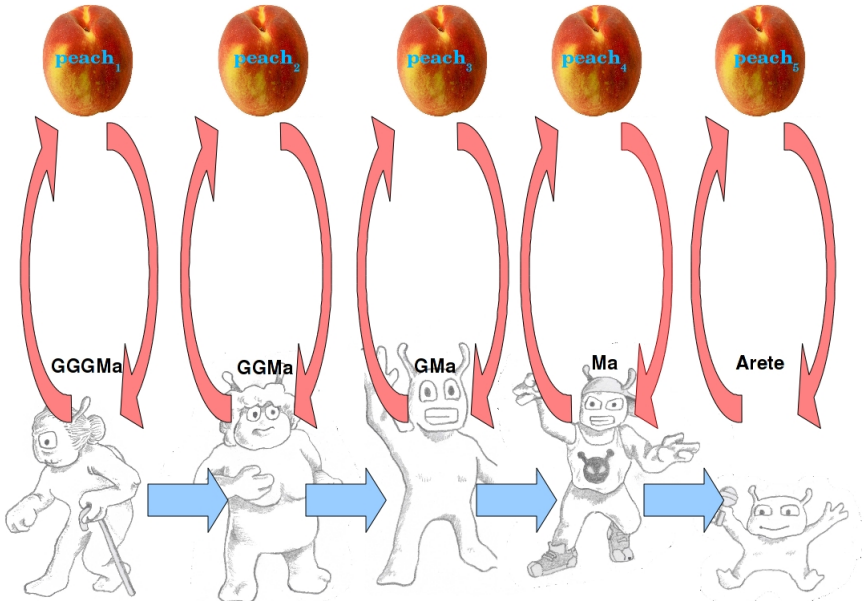

Figure 3: Many Earths, Many Peaches

> In general, the n^{th} member of the family line has been abducted, by the most bizarre of coincidences, to a different planet $Earth_n$. They have all fed on the peachy nutritious things around (peach, $peach_2$, ..., $peach_n$) and this has helped them be better off than conspecifics not endowed with M. This has accounted for M's selection, to the extent that, nowadays, M has become fixated in Arete's population (*cf.* figure 3.)

ETIOLOGICAL FUNCTION warrants the attribution to M of the function to indicate some properties. Take, for instance, the property F: *Having peachy-looks and being nutritious for the likes of Arete*[1].

- EF1: M has indeed indicated the instantiation of F around its possessor in a sufficient number of Arete's ancestors. It has indicated it in *all* of them, actually: for all of her ancestors, $P(on|F) > P(on)$ and, moreover, this conditional probability is causally grounded, for each ancestor *n*, on $peaches_n$.
- EF2: *Ex-hypothesi*, M's fitness contribution as seen from the perspective of indicating Fs has been positive (that is, $FC_F > 0$) for Arete's ancestors, and this is part of an explanation of the fact that Arete has M.
- EF3: There is a causal explanation for (a sufficient number of) the conditional probabilities in IP_F. The main element of the indication profile, $P(on|F)$, as we have said, is grounded on $peaches_n$: it is these fruits that ground the correlation between *Having peachy-looks* (which is M's input, remember) and *Having peachy-looks and being nutritious for the likes of Arete*. We may further assume that the overall frequency of peachy things which are not nutritious, and that of nutritious, non-peachy things that M's possessor manages to eat, is comparatively low. This grounds the rest of IP_F.

1 If there are such conjunctive properties. This is the kind of hybrid, input-output properties that Agar (1993) defends as content attributions for the likes of Democritus and Arete.

- EF4: The positive fitness value of hits derives, precisely, from the fact that they result in the ingestion of nutrients; now, every peachy thing that is nutritious is, unsurprisingly, nutritious. Whatever grounds P(on|F), then, grounds w_{11}.

In conclusion, M has the function of indicating instantiations of the property *Having peachy-looks and being nutritious for the likes of Arete* around its possessor. As a consequence, BETTER DRETSKE warrants the attribution of the content *There is something with peachy-looks and nutritious for the likes of Arete around S.*

On the other hand, there is no specialised homeostatic mechanism that helps explain the frequent coinstantiation of the properties that ground IP_F and FM_F across generations. Nothing, but chance, explains that. That ingredient, normally present wherever there is selection for a mechanism such as M, is absent here. Arete's M has, therefore, not managed to zero in on any particular real kind. It is, according to THERE IS AN F AROUND, contentless.

So if, as I have been urging, solving the Indeterminacy Problem for simple states depends on relying on an HPC that ties together the causal explanation across generations of the existence of an indicator, then we have to live with the fact that some states that have the function to indicate are contentless. I have argued in 1.6 that this should not be motive of much trepidation: attributing content is a compressed way of claiming that there is a certain natural structure such that the existence of a mental state enjoys a certain, quite concrete kind of explanation based on the former. It may be that function attributions rely on less specific explanations than content attributions; a greater number of causal paths may lead to selection for an indicator than lead to contentful states. In those (most likely never instantiated) cases it is not implausible to claim that we have enough with the pre-semantic notion of *indication*, and content attributions are overattributions.

The reader may, on the other hand, cherish the insight that the contentful states are all and only the states that have the function to indicate, and she may wish to strengthen the notion of etiological function so that these two classes are again coextensional. I'm going to discuss now whether this is a sensible project.

2.1.1 *Strengthening* ETIOLOGICAL FUNCTION.

Could the strict teleosemanticist, in light of the foregoing discussion, strengthen her definition of function, adding a condition EF5?

EF5: The fact that EF3 and EF4 remain true for (a sufficient number of) S's ancestors is causally grounded.

First of all, it is unclear that she *should*. It is natural to think that M in MANY EARTHS, MANY PEACHES has kept on existing *because* of the indication relations it has established with its environment. And this history, it appears, is the kind of thing that warrants etiological function attributions.

In fact, I suspect that even conditions EF3 and EF4 are spurious as requisites for a *function* attribution to be warranted: rather, they are there because ETIOLOGICAL FUNCTION is a building block for the construction of a content-attribution recipe. And, as SELECTION FROM FITNESS CONTRIBUTION states, a naturalistic account of the

biological function of a mechanism should only rely on whether that mechanism does something that makes its possessor reproduce more than competing conspecifics. For instance, the much more parsimonious INDICATION* and ETIOLOGICAL FUNCTION* below also abide by said tenet.

INDICATION*: A mechanism M's going *on* indicates instantiations of a property F around S iff

I1: $P(F|on) > P(F)$

ETIOLOGICAL FUNCTION*: A mechanism M in a subject S has the function of indicating the instantiation of F around S iff

EF1: According to INDICATION*, M has indicated the instantiation of F around its possessor in (a sufficient number of) S's ancestors.

EF2: M's fitness contribution as seen from the perspective of indicating Fs is positive (that is, $FC_F > 0$) for S's ancestors, and this is part of an explanation of the fact that S has M.

But BETTER DRETSKE would yield wildly implausible results using these two definitions: for example, any random, and randomly beneficial, correlation between the indicator* being *on* and the instantiation of a property anywhere would count as providing a content for the indicator*. So, if the ancestors of Democritus's token of M's being *on* has correlated, by a bizarre coincidence, with gamma outbursts near Alfa Centauri, and this has also coincided, randomly, with these ancestors being better off after M fired, INDICATION* and ETIOLOGICAL FUNCTION* would conspire to warrant a content attribution to M's being *on* of *There has been a gamma outburst near Alfa Centauri*. On the other hand, THERE IS AN F AROUND could use INDICATION* and ETIOLOGICAL FUNCTION* without any changes in what is deemed or not contentful

In summary: maybe the notion of function used, or presupposed, when formulating teleosemantics is unduly strengthened in a way that suits a theory of content but goes beyond what should be countenanced by a theory of functions[2]. All the more so if we throw EF5 in.

Moreover, adding a condition EF5 does not solve the Indeterminacy Problem. Admittedly, such a definition ETIOLOGICAL FUNCTION** (i.e., ETIOLOGICAL FUNCTION + EF5), as we wish, does not warrant a function attribution to M in *Many Earths, Many Peaches*: EF5 is not met. So far so good, but, whenever there is an HPC there to comply with EF5, we still have the whole indeterminacy there for us. Even on Earth, where peaches play the role that EF5 demands, it is still the case that ETIOLOGICAL FUNCTION** warrants a function attribution to M of indicating *Having peachy-looks*, *Being nutritious for Arete*, *et* a great many *caetera*. Indeed, this was the conclusion of section 1.3.

The only way to fix ETIOLOGICAL FUNCTION to avoid indeterminacy, in the direction I have been advocating, is to rule that only HPCs are such that an M has the function of indicating them, thus:

ETIOLOGICAL FUNCTION***: A mechanism M in a subject S has the function of indicating the instantiation of F around S iff

2 I think Price (1998) is a particularly vivid case of unduly strengthening the notion of function on the way to solving a problem in semantics. See footnote 3.

EF1: According to INDICATION, M has indicated the instantiation of F around its possessor in (a sufficient number of) S's ancestors.

EF2: M's fitness contribution as seen from the perspective of indicating Fs is positive (that is, $FC_F > 0$) for S's ancestors, and this is part of an explanation of the fact that S has M.

EF3: There is a causal explanation for (a sufficient number of) the conditional probabilities in IP_F.

EF4: There is a causal explanation for (a sufficient number of) the conditional probabilities in FM_F.

EF5: The fact that EF3 and ef4 remain true for (a sufficient number of) S's ancestors is causally grounded.

EF6: F is the natural kind individuated by the properties together with the homeostatic mechanism that explain that EF2 to EF5 are met.

EF6 should be spelt out in detail, as we have done for the construction of THERE IS AN F AROUND. If this is done wisely, BETTER DRETSKE*** (i. e., BETTER DRETSKE after substituting ETIOLOGICAL FUNCTION with ETIOLOGICAL FUNCTION***) should end up extensionally equivalent to THERE IS AN F AROUND.

There remain two questions about such a view: one is whether this is at all a sensible analysis of the notion of having the function to indicate; whether this is salvaging the letter of the teleosemantic proposal at the cost of its spirit. Intuitively, it seems that many more things in the world have a function to indicate that are countenanced by ETIOLOGICAL FUNCTION***. In particular, the most sophisticated etiological developments of the notion of etiological function (Millikan's among them) are less restrictive. On the other hand, at least one prominent account is as restrictive as ETIOLOGICAL FUNCTION***: Price (1998). Price explicitly states that her theory is intended as a solution of the Indeterminacy Problem in teleosemantics. Now, I think it is dangerous, as a general methodology, to try to solve problems in one theory (the theory of content) by tampering with the reducing theory (the theory of functions). In this case, as I have explained, the tampering has the unwelcome consequence that less things are deemed to have the function to indicate than what is plausible. It is a more satisfactory solution overall to identify contents with the structures that enable the appearance of functions to indicate, as I have explained[3].

3 There are other problems with Price's account of functions. I cannot provide a close examination of her theory, but there are at least the following two issues:

- The Abstractness Condition (*op. cit.* p. 67f) states that the explanation of the existence of the functional item must not depend on *specific* features of the design of said items. This notion of explanation is, as far as I can see, intentional. Price provides no principled, non-intentional way to ascertain whether the condition is met. But in the context of the program of naturalising semantics, this is not an acceptable situation.
- It is extensionally inadequate. Price's final analysis of function-talk is:

  Where d is a device and G is an activity, d has the function to do G iff there is some type of item E; there is some item e which belongs to E; there is some type of device D, to which d belongs, and which consists of devices produced by E-type devices in some manner M; and there is some activity F, such that:

  1. This d was produced in manner M by e;
  2. D-type devices have done G;

2.2 THE SELECTED-EFFECTS RESTRICTION AND CONSUMER SEMANTICS

It is features about the causal explanation of the existence of contentful states, over and above those that fix the function of the mechanism that produces them, that have helped us solve the Indeterminacy Problem for the content of simple states: the recurrence of a property cluster, which some homeostatic mechanism keeps in place, helps explain such existence while at the same time individuating a natural kind. The causal requisites of function-attribution are less strict.

In many senses etiosemantics is very congenial with traditional teleosemantics: the explanations of the existence of M I have appealed to are of the kind standardly used to ground function attributions, although they go beyond those needed to ground such functions. I will now discuss in more detail other respects in which the proposal at issue fares, I think, better than its competitors. In particular, better than the very influential family of views frequently discussed under the name of *consumer semantics* and, specially, Millikan's account.

I will start with an important objection to Dretskean teleosemantics as codified in BETTER DRETSKE: the *Selected Effects* restriction. Functions are what devices are supposed to *do* -as opposed to what devices are supposed to *be done to*. As Neander puts it,

> biological proper functions are *effects* for which traits were selected by natural selection. Neander (1991, p. 168, my emphasis)

This restriction makes it difficult -though not impossible- to account for the very plausible idea that the content of some mental states has partly to do with whatever has *caused* it in its ancestors. Given that, as Neander points out, functions are selected effects, strict teleosemanticists cannot directly help themselves to the input pattern of a mental state in order to fix its content; even if the input pattern has played a crucial explanatory role. Thus, Millikan:

> A problem with Dretske's view is that it is hard to see how it could be the function of any biological device literally

3. The fact that e produced this d by M is explained partly by the fact that by doing G, D-type devices have helped to bring it about that e or other E-type items which are ancestors of e did some further thing F;
4. There is nothing which D-type devices did in the past which would provide a more immediate explanation of e producing this d;
5. G is characterised in terms of some effect which D-type devices were able to bring about by themselves;
6. The truth of the explanation referred to in 3 does not depend on specific features of the design of those D-type devices, or of any other devices which worked in conjuction with those devices to bring it about that e or e's ancestors did F.Price (1998, p. 68)

Consider the following case: someone starts sweating, the sweat wets his shirt and this makes him uncomfortable, which, in turn, causes further sweat. And take:

- d: a drop of sweat.
- D-type: drops of sweat.
- G: wetting shirts
- e: a sensation of social uncomfortability
- M: nervous sweating
- F: Increase, or perpetuate.

In this case, certain drops of sweat have the function of wetting shirts.

> to *effect* the production of one of his "indicators." To do so, the device would have to *effect* that certain statistics should hold. Millikan (1993, p. 129)

As I have said above, I think it is correct to insist in the fact that functions cannot be causally upstream from the functional device. This is an important requisite for the philosopical theory of functions. I have also said that we should not feel constrained by this requisite when building our theory of *content*. If the best theory of content involves, as I think it does, mechanisms whose selection has been, in part, explained by their Indication Profile, so be it. We should stop talking of functions in the reduction base, and rather talk of pseudo-functions, and move on.

Anyway, if we stick to functions as opposed to pseudo-functions, we'll need to say that the function of a mental representation will have to do solely with whatever it is supposed to cause downstream in the cognitive system of which it is part. That's how we get to another important insight of traditional teleosemantics; the one in virtue of which it is also called *consumer semantics*: it is facts about the way the output of the representation (*i. e.*, whatever it causes) is utilised (*consumed*) by mechanisms causally downstream which fix (in a way that varies according to each theory) the content of the representation.

Some summaries and expositions of consumer semantics make (what I take to be) an error at this point[4]. Take, for example,

> On the teleosemantic approach, content depends on how consumer mechanisms interpret representations. It depends on the behavioural output, not the informational input. The content is that condition under which the resulting behaviour would be appropriate, whether or not the actual circumstances that caused the representation are of that type. Macdonald and Papineau (2006, p. 6)

It is simply not true that consumer semantics are committed to identifying content with the "condition under which the resulting behaviour would be appropriate"; at least because this would be a terrible theory. Take, for instance, MacDonald and Papineau's example of a content attribution according to this paradigm:

> [A mental representation] is a snake representation because it makes you behave in a way appropriate to snakes, given your biological ends. And this will remain the case even if you are pretty bad at recognising snakes. The production mechanism for this representation may be triggered by toy snakes, by other slithery animals, indeed by the slightest hint of a slither, yet the representation will still stand for snake, if it is specifically snake-appropriate behaviour that it prompts. (*ibid*.)

It may well be that a snake-representation remains so "even if you are pretty bad at recognising snakes", but, under this view of consumer semantics, it does *not* remain a snake-representation if you are pretty bad at *displaying snake-appropriate behaviour*. Now, surely, less-than-perfect snake-appropriate behaviour (behaviour that does not distinguish between, say, snakes and lizards) can be fitness improving in many cases.

4 This may be simply necessary to fulfil their roles as summaries and expositions, which calls for a sacrifice in rigour to improve clarity.

Moreover: there are many situations in which the most appropriate behaviour is not snake-selective in the least. When a mouse flees it does not do it in any particularly anti-slithery way. But, under McDonalds and Papineau's version of consumer semantics, one needs such selectivity to snakes in the consumer end of the representation to credit it with snake-involving content. This would be, as I say, a very uncomfortable result.

But consumer semantics can choose other, more reasonable ways in which the way the representation is consumed helps fix its content. For example, in Millikan's biosemantics, contents are to be identified with the conditions that held in the occasions in which past responses to the output of the representations helped explain the selection of the representation's producer. Let see the process at work in Democritus's example:

1. A certain mechanism (M, in our example) produces a couple of representations (M's being *on* and M's being *off*).

2. M's being *on* is consumed in the following way: it causes its possessor to protract its tongue. This has been useful in a number of occasions because, at the right end of the tongue, there was nutritious stuff. As a result, M has been selected for. Part of what makes M useful, also, is that it is *off* most of the time when there is no food around. This avoids idle resource expenditure in protracting the tongue when there is nothing good to catch.

3. M's being *on* and its being *off*, in the relevant situations -*i. e.*, when there has been selection for M- correspond to (or, in Millikanian terminology, *map onto*) the conditions of *There being frog food around* M*'s possessor* and *There not being food around* M*'s possessor*, respectively. These are to be considered the contents of the two states.

> [T]he systems that use, that respond to, the frog's fly detector signals, don't care at all whether these correspond to anything black or ambient or specklike, but only whether they correspond to frog food. (...) So the firing means frog food. Millikan (1991, p. 163)

I have stated in 1.3.2 that Millikan's biosemantics is subject to the Output Problem. We can now see why: in 2. above, after "This has been useful in a number of occasions because, at the right end of the tongue, there was" we can write many things apart from "nutritious stuff". In fact, what is important is that at the right end of the tongue there is *non-poisonous nutritious stuff* -nutritious, but otherwise toxic stuff not being what makes tongue protracting useful- so this is what a consumer-semanticist should settle in for as content. Or should it be *non-poisonous frog food such that there is no frog-predator near it*? Etc.

Millikan's account, in cases such as Democritus's, seems to have to settle in for the highest content attribution; maybe something along the lines of *There is something good for frog digestive systems around*. This attribution seems artificial or, at any rate, more artificial than the content etiosemantics attributes in this case: *There is a fly around*.

Millikan has mounted a sophisticated defense against this charge, involving what she has recently been calling *local natural information* (cf., *e. g.*, Millikan (2004, 2009)). Before discussing this defense, I will

point out one important difference in the predictions that etiosemantics and biosemantics make.

2.2.1 *Millikan's Normal Conditions.*

We have seen the way in which Millikan abides by the selected-effects restriction while still honouring the importance of external conditions, the ones that cause mental states to fire, in fixing the content of representations:

> The content of the descriptive sign is not determined by the tasks its consumer performs. It is determined by what the sign needs to correspond to if the consumer is to perform its tasks in its normal way. The producer's job is merely to make a sign that corresponds in the right way to a world affair. If it does this in its normal way, by its normal mechanisms, the intentional sign it makes will also be a natural sign. Millikan (2002, p. 79f)

A mechanism such as Arete's M in ARETE AND THE PEACH TREE has a function, *i. e.*, a collection of selected effects -roughly, making Arete grab and eat the things that make it fire. That is, the consumer of M, C, is some relevant part of Arete's motor control engine, whatever it is. We may assume that C contributes to Arete's fitness whenever the things it makes Arete grab and eat are nutritious for her. This is enough for C's "performing its tasks". As with Democritus, the descriptive sign in this example is M's going *on*, and M is the producer of these signs.

So, according to Millikan, it is enough for M to fulfil its function that it goes on whenever the grabbing and eating will secure *nutritious stuff for* M*'s possessor*. Under this perspective, in ARETE AND THE PEACH TREE, Ma (the one that lives and thrives in Twin-Earth, out of twin-peaches) has an M that performs its function impeccably: it contributes to Ma's fitness by providing correct advice about what to grab and eat, and what not to.

But M's function is not the only source of normativity for M's performance. The *way* in which it fulfils its function is also subject to appraisal: there is a privileged, *normal* way for M to work. This normal way[5] must be cashed out as involving the conditions in which, historically, M was when it performed its function (*cf.* Millikan (1984, p. 33f)) If so, M does not contribute to Ma's fitness in a normal way: if it has signalled *nutritious stuff for* M*'s possessor*, it has historically done so by relying on the causally-grounded correlation between peachy-looks and nutritiousness facilitated by peaches. That very causal grounds are not present in Ma's case; so M is not a natural sign of nutritiousness in the abnormal conditions present in Cheap Earth. But in normal conditions it is such a natural sign. This effects the reconciliation between the selected-effects restriction and the fact that representations are, normally, natural signs.

One must try to avoid being lulled by vocabulary such as "normal" or "selection". Millikan's proposal is made in the context of a naturalising effort. In that context, such vocabulary is admissible only if it can be cashed out in clearly naturalistic terms. "Selection" is, at bottom, differential reproduction in some natural context or other. "Normal" is whatever feature was present during selection. Now, we can easily envisage cases in which normal conditions as defined are not, what

5 In older writings by Millikan it used to be "Normal", capitalised.

we would intuitively call, *normal*. In those conditions, Millikan's contention that the intentional sign will be a natural sign is wrong. MANY EARTHS, MANY PEACHES is, precisely, one such case: the conditions prevailing during selection of M (that is, the *normal conditions*) do not allow M to zero in on any particular natural sign of nutritiousness. What natural signs provide in everyday selection cases, an overwhelming good luck provides in this particular case. Millikan's biosemantics yields the prediction that M's being *on* has a content along the lines of *There is nutritious stuff for the likes of Arete around here*; but, if this is so, M's being *on* has a nutritiousness-involving content without M's being *on* being a natural sign of nutritiousness. This is not a problem in general for contentful states; we have a great many contentful mental states which are no natural sign of their content -beliefs about impossibilities are an extreme case. But, I think, it is a problem for these very simple states which, intuitively, have content in virtue of their tracking substances around them.

I have stated in 2.1 that content is doing no real job in these cases, and that it is explanatory enough to say that M has the function to indicate that property -or, we might add now, the selected-effects cognate that Millikan prefers. We have in the present discussion further reasons to think that this is so: Millikan rescues natural signhood as a consequence of the mechanism of selection in everyday cases. But this puts things upside down: what happens in fact is that environments are, normally, stable enough to allow for selection for some mechanisms which rely on this stability. Content should be identified with this selection-enabling stable features. If not, in the abnormal -but nomologically possible- situations in with selection takes place without environmental constancy, we are forced to provide content attributions which are uncalled for.

THE USE OF THOUGHT EXPERIMENTS. Millikan has repeatedly (*e. g.*, in Millikan (1989a) and Millikan (1989b)) voiced her disapproval of thought-experiment-driven philosophy. She does not intend her elucidations of function, sense, content and the like to be analyses of the common-sense concepts behind those terms. Rather, they are to be appraised solely on the basis of their theoretical fruitfulness. If a theoretically fruitful concept conflicts with some of our intuitions about the applicability of its common-sense counterpart, so much the worse for the latter. So, a thought experiment cannot prove wrong Millikan's definitions, which are postulations, not analyses. If this is correct, MANY EARTHS, MANY PEACHES has no force against Millikan's attributions of function and content.

I am sympathetic with this qualified scepticism about thought experiments and, like most everyone else, agree that the right attitude to take is one that strives for a reflective equilibrium between intuition and theory. There are a couple of things that may be rejoined in this connection, though. First, the role of the thought experiments presented in this chapter is only to make vivid a theoretical proposal: that content is to be thought as fixed by natural structures (such as HPCs) that enable selection for indicators. Thought experiments can be used to show that, although it may be assumed that Millikan's biosemantics adheres to that proposal, it actually does not. Second, a theory's conforming with our intuitions in the evaluation of one thought experiment or other, we may agree with Millikan, is not a terribly important *datum* in favour of said theory. But it surely is not *against* the theory that it explains some

of our intuitions. All in all, if we have to choose among two theories whose predictions coincide among them and with common sense in all everyday scenarios[6] and differ in their prediction about some *recherché* scenario, I submit, it is rational to accept the theory that tracks common sense in that very one scenario.

2.2.2 *Biosemantics and the Output Problem*

In a recent summary of Millikan's views, Millikan (2009), there is a brief (but, as far as I am aware, the most explicit) discussion of the Output Problem:

> Taking for her example the female-hoverfly detector in a male hoverfly's visual system, Karen Neander (1995) has objected that among the external conditions needed for the detector's consumers to perform all their functions are that the female is fertile and that she won't be eaten before she reproduces, hence that on the biosemantic theory these facts about the female must be part of what is represented by the detector in the male's visual system. What this overlooks, however, is that an intentional icon [for our purposes, this is interchangeable with what I have been calling a simple contentful state - MM] must also have a producer and that it must be a function of the producer to make an icon that corresponds to the condition it represents. If the producer has a function there must be a normal mechanism by which it performs that function. This, however, would require the male hoverfly's visual systems to be sensitive to natural signs of fertility in female hoverflies and of liability not to be eaten. But on no theory of information, certainly not on the theory of local natural information, does the male hoverfly use or even encounter any such natural information. Millikan (2009)

This discussion presupposes the following setting: male hoverflies have a detector (we will call it M) that fires when a shadow of a certain shape and at a certain speed crosses the hoverfly's retina. M's going *on* causes the hoverfly to dart in a certain direction, calculated from the speed and angle of the shadow, which in a sufficient number of times helps the hoverfly reach a fertile female to mate with. The case is described in Millikan (1990), Millikan (1993).

Millikan's argument seems to be the following.

1. If *There is a fertile female hoverfly around* is to be warranted by biosemantics as a content-attribution to the state M*'s being on* of a male hoverfly, the state's producer M should have as a function to produce a state that corresponds to the condition of being a fertile female hoverfly.

2. If M has such a function, there must be a normal mechanism for its fulfilment.

3. Such normal mechanism must involve M's sensitivity to natural signs of fertility in female hoverflies.

6 I don't think this is the case with consumer- and etiosemantics, but let that pass.

4. But the male hoverfly does not encounter information upon which to build such a sensitivity.

5. Hence, the content-attribution in 1 is not warranted.

I have just argued against 3 above: it is not true that normal mechanisms *must* involve sensitivity to natural signs. If there are *no* such natural signs, normal mechanisms cannot involve sensitivity to them, and I have shown how etiological functions of the relevant kind may emerge -to be sure, only in bizarre, unlikely cases- in the absence of natural signs. Anyway, this is a moot point here: even if 3 *need* not be the case, it *is* surely the case as a matter of fact in actual hoverflies.

The real difficulty, it seems to me, comes with 4. It is, to begin with, not true that "on no theory of information" does the male overfly encounter information about fertility in female hoverflies. Take INDICATION: in many cases we may perfectly well have that

11: $P(\mathtt{Fertile}|\mathtt{on}) > P(\mathtt{Fertile})$ and

12: The difference in probabilities in 11 is causally grounded.

That is, M indicates instantiations of the property of *Being a fertile female hoverfly*[7], and this indication relation is causally grounded, even if M is unable to distinguish at all between fertile and infertile hoverflies. *I. e.*, even if $P(\mathtt{on}|\mathtt{Fertile}) = P(\mathtt{on}|\mathtt{Infertile})$. This is because the percentage of female hoverflies that are fertile stays approximately constant, for reasons having to do (I hypothesise) with the rate of genetic mutation and environmental conditions leading to infertility, and which provide the causal grounding for the relevant probabilities. In such a case, M indicates instantiations of the property of *Being a fertile female hoverfly* without needing to exploit a natural sign that is specific to fertile (as opposed to infertile) females[8].

What about Millikan's own theory of *local natural information*? Although, in the passage quoted, she explicitly denies that M carries local natural information about fertile female hoverflies, it is unclear that this follows from the characterisation she has made (in Millikan (2004, chapter 3) and Millikan (2007)) of this notion:

> [A sign carrying natural local information] is one that corresponds to its represented in the same way, and for the same reason, that other signs of the same recurrent type correspond to theirs, and where there is a reason why examples of this kind of correspondence (with the same kind of cause) tend to spread from one location into nearby space-time locations. Millikan (2007, p. 453)

If this is a strict definition, the matter is easy to settle: M *does* carry natural local information about fertile female hoverflies. There are causal reasons -mutation rates, environmental conditions; see above- why a sign that loosely corresponds to female hoverflies also loosely corresponds to fertile female hoverflies, and a reason why this kind of correspondence tends to spread -the rate of fertile hoverflies gets

7 Or, in Shea (2007)'s terms, M carries *correlational information* about it. See next section.

8 By, the way, it is possible that it also indicates the property of *Being an infertile hoverfly*. This is not a problem for Millikan: such property is, surely, not the one the consumer needs and therefore is out of the question as a candidate for content. It is not a problem for me either: it is one of the properties the specialised homeostatic mechanism that fixes the content brings along for the ride. See below for my take on the hoverfly case.

copied with hoverfly reproduction, relevant environmental conditions stay put, etc.

In an earlier discussion Millikan claims that the reason why M carries local natural information about female-hoverflyhood but not fertile-female-hoverflyhood is that,

> The domain in which the hoverfly operates is one in which the chance that the shadow crossing its retina, assuming that it is of a female hoverfly, is also of a fertile female not about to be eaten is no higher than the chance of any arbitrary female hoverfly being fertile and not about to be eaten. By contrast, assuming that it is the shadow of a hoverfly, the chance of the shadow being that of a female is considerably higher than the chance of an arbitrary hoverfly being female. Millikan (2004, p. 85f)

That is, on the one hand,

$$\begin{aligned} P(\mathit{fertile\ female\ hoverfly} | M \text{ is on} \wedge \mathit{female\ hoverfly}) \quad &\leqslant \\ P(\mathit{fertile\ female\ hoverfly} | \mathit{female\ hoverfly}) & \end{aligned}$$

while, on the other hand,

$$\begin{aligned} P(\mathit{female\ hoverfly} | M \text{ is on} \wedge \mathit{hoverfly}) \quad &> \\ P(\mathit{female\ hoverfly} | \mathit{hoverfly}) & \end{aligned}$$

The theoretical reason behind the change of conditions in the probabilities above and below is not perfectly clear. For example, it is also true that

$$\begin{aligned} P(\mathit{fertile\ female\ hoverfly} | M \text{ is on} \wedge \mathit{hoverfly}) \quad &> \\ P(\mathit{fertile\ female\ hoverfly} | \mathit{hoverfly}) & \end{aligned}$$

which, paraphrasing Millikan, means that

> ... By contrast, assuming that it is the shadow of a hoverfly, the chance of the shadow being that of a fertile female is considerably higher than the chance of an arbitrary hoverfly being a fertile female.

Again, this is so because of the fact that the rate of fertile females hoverflies in the female hoverfly population is sufficiently large, and a number of causal processes ensure that this remains so. But this is, if we are to judge by the passage just quoted and for reasons that are quite unclear, irrelevant to M's carrying local information about fertile-female-hoverflyhood.

Moreover, in the same passage, Millikan defends that M carries local natural information about hoverflyhood because

> the chance of the shadow crossing [the hoverfly's] retina being that of a hoverfly rather than that of some other small particle of matter is also very much raised. (*ibid*.)

Which, I take it, can be rendered as

$$\begin{aligned} &P(\mathit{hoverfly}|M\ \mathit{is\ on} \wedge \mathit{some\ particle}) \quad > \\ &\qquad P(\mathit{hoverfly}|\mathit{some\ particle}) \end{aligned}$$

But, again, it is also true that

$$\begin{aligned} &P(\mathit{fertile\ female\ hoverfly}|M\ \mathit{is\ on} \wedge \mathit{some\ particle}) \quad > \\ &\qquad P(\mathit{fertile\ female\ hoverfly}|\mathit{some\ particle}) \end{aligned}$$

That is, paraphrasing Millikan again,

> the chance of the shadow crossing [the hoverfly's] retina being that of a female fertile hoverfly rather than that of some other small particle of matter is also very much raised.

All in all, there is no principled reason to deny that M carries local natural information about fertile-female-flyhood. And, without such a reason, Millikan's biosemantics is still subject to the Output Problem.

Millikan (in personal communication) has suggested that this discussion shows, indeed, that she should accept *There is a fertile female hoverfly around* as the right content attribution to M's being *on*. But that is as far as we need to get: further attempts to push the biosemantic content attributions towards the highest attribution will involve properties which M does not carry natural information about; say, *Being a fertile female hoverfly such that it won't be eaten before it reproduces*. Surely that's not the kind of thing M can carry information about?

In fact, I think it is a virtue both of Millikan's theory of local natural information and of INDICATION that, according to them, mental states do carry information about (indicate) these properties. Again, the flying stuff that causes the right kind of shadows on the hoverfly's retina is, more often than not, a female hoverfly, which are, more often than not, fertile, which in turn are, more often than not, lucky enough not to be eaten before reproducing -and this not by chance, either: the density of predators is what it is, and the homeostatic properties of the ecosystem will ensure that this remains so. Of course, if you nest enough *[more often than not, for a reason]* operators, you will reach a property that M does not indicate (carry information about). But the highest property such that M indicates it will already be too high to be a natural content-candidate.

It is maybe useful to see how etiosemantics deals with the hoverfly case. Part of the interest of the case has to do with the *rule* the male hoverfly follows to calculate its response to the retinal shadow. In this connection, Millikan has defended that biosemantics offers a response to Kripkenstenian sceptical considerations. I will have something to say about Millikan's discussion of this feature of the case in 4.5 but, for the purposes at hand it is enough if we concentrate on a content attribution to M's being on of the kind *There is an F around*. We have seen that biosemantics is forced to put very high properties in place of the F. What about etiosemantics?

We know that M indicates, better or worse, a number of properties: *Being a flying thing*, *Being a female hoverfly*, *Being a fertile female hoverfly that will not be eaten before reproducing*, etc. M has an Indication Profile and a Fitness Matrix for each of these properties; and each such pair of Indication Profile and a Fitness Matrix allows to calculate a (the same) Fitness Contribution according to which ETIOLOGICAL FUNCTION

warrants attributions to M of the function of indicating instantiations of all of those properties. We may, in the way shown in 1.4.1 and 1.4.2, find out which properties must be frequently coinstantiated around M's possessor. We will end up with a list of properties including:

- Being a body of more or less such and such a shape darting at more or less such and such a speed
- Being a suitable mating partner for a male hoverfly

And many others. There is a specialised homeostatic mechanism that keeps most of these properties together in an spatiotemporal region (a, t) that overlaps sufficiently with the region in which M gets selected for: hoverfly reproduction plus the sex-determination process that results in a female, plus enabling conditions such as an atmosphere, etc. The smallest number of properties that such a specialised homeostatic mechanism keeps together in (a, t), together with the mechanism itself, individuate an HPC that may be plausibly taken to be identical to the real kind *female hoverfly*. The content of M is, thus, *There is a female hoverfly around*.

There is a standard rejoinder to objections such as the Output Problem from friends of teleosemantics: it is, after all, a good thing that such simple creatures as frogs and hoverflies are predicted to have mental states with slightly unfocused contents. Perfectly determined contents are best reserved for sophisticated cognisers such as human beings. In particular, according to Millikan, it is open to the theorist to suppose, plausibly, that the consumer systems in human brains are extremely picky, and that satisfying their needs will require representations to map onto much more precise states of affairs. I would like to say something about this rejoinder.

First of all, a minor point. This rejoinder looks like an afterthought: teleosemantists have stumbled upon this family of problems, and given the difficulty of providing what, after Schiffer (1996), we may call a *happy face solution* to the problem, some of them have settled in for the "unfocused is fine" reply. Etiosemantics offers a true happy face solution to the problem: the content of frogs' and hoverflies' mental states involve, as we originally wished, flies and hoverflies. This is maybe a reason to rethink the afterthought.

A second, maybe more important worry is that the content biosemantics predicts for the male hoverfly's mental state M is not really unfocused or indeterminate. The content, very determinately, involves the highest property such that M's being *on* carries natural local information about it, be it *Being a fertile female hoverfly*, *Being a fertile female hoverfly that will not die before reproducing* or whatever. The problem is not the indeterminacy, but the implausibility of the content attribution. Etiosemantic attributions are, *prima facie*, more appealing.

This is obviously not a knockdown argument against biosemantics: our intuitions regarding the contents that frogs and hoverflies entertain are shaky at best, and we may well sacrifice a measure of plausibility in these, the shady corners of the theory if it means better fit with the paradigmatic data: human contents. A third problem with the Millikanian rejoinder comes at this point: that the human consumer-systems are picky enough to solve the Output Problem is an interesting hypothesis, but one that has not been developed in anything like the necessary amount of detail even in Millikan's writings. As matters stand, It may not be unwise, if possible, to design a theory that provides

determinate contents in these very basic stages, which may then work as building blocks for more sophisticated contents. I submit etiosemantics may be such a theory.

2.3 SHEA'S INFOTEL SEMANTICS

Recently, Shea (2007) has advocated a Dretskean theory of content as a solution to an objection by Godfrey-Smith (1996) to traditional teleosemantics. In this section I will, first, briefly present the objection, along with Shea's solution; after that, I will show that Shea's *Infotel semantics* falls prey to the Indeterminacy Problem for exactly the same reasons that other broadly Dretskean proposals -such as the one presented in 1.2- fail. I will finally show that etiosemantics solves the Shea/Godfrey-Smith objection at least as well as infotel semantics and is, therefore, a better package deal.

2.3.1 *The Behaviour-Explanation Objection*

One of the chief uses to which we put content attributions is the explanation of successful behaviour. Thus, *e. g.*, our successful goings-to-the-fridge are explained by our there's-a-beer-in-the-fridge doxastic states -Shea (2007, p. 410f); see also the introductory remarks to chap. 3 in Dretske (1988).

But, the objection goes, an explanation of a piece of behaviour according to which such behaviour is caused by a representation with thus-and-so a content is substantive only if the content of representations is not, in its turn, fixed by appealing to the success conditions of the behaviours it tends to cause. That would be a very thin explanation, if not outright circular.

And it is precisely, according to Shea, the teleosemantic content-fixing strategy:

TELEO: In the past R caused a consumer subsystem to behave in a way that contributed systematically to survival and reproduction only if R truly represented that C. Shea (2007, p. 416)

According to Shea, the traditional teleosemanticist wishes to defend that the left-hand side is constitutive of R's truly representing that C. Take now the following schema of a content-based explanation of successful behaviour.

BEHAVIOUR EXPLANATION: The piece of behaviour B has contributed to survival and reproduction because it was caused by a representation with the content C.

Plugging TELEO into BEHAVIOUR EXPLANATION, we get

BEHAVIOUR EXPLANATION - TELEO: The piece of behaviour B has contributed to survival and reproduction because it was caused by a representation R such that, in the past, things of the same type caused a consumer subsystem to behave in a way that contributed systematically to survival and reproduction.

There are a couple of minor wrinkles in the derivation of the very thin BEHAVIOUR EXPLANATION - TELEO from BEHAVIOUR EXPLANATION and TELEO: first, it is unclear *whose* survival and reproduction

is relevant in BEHAVIOUR EXPLANATION. The "chief use" of content attributions appealed to above is in order to explain the behaviour of cognitive subjects such as persons or animals. It is much less common to use content attributions to explain the *behaviour* of their cognitive subsystems. If so, it is persons' and animals' survival and reproduction that is relevant in BEHAVIOUR EXPLANATION, and TELEO (which talks about survival, reproduction and behaviour of cognitive subsystems) should be modified before plugging, maybe by working out the not entirely straightforward relation existing between animal-survival and subsystem-survival.

Second, teleosemantics is not such a thin explainer as Shea appears to suggest. According to traditional teleosemantics, there are other facts constitutive of R's having the content C that are explanatorily relevant, apart from the fact appealed to in TELEO (*i. e.*, that, in the past, a number of behaviours have contributed to survival as a result of being caused by R). In particular, the cognitive subsystem whose behaviour we are in the business of explaining has a *concrete* function in the mental economy of its possessor. It is this concrete function that fixes the behaviour of R, not just some general usefulness of its doings. Teleosemantics, that is, can at least provide the following[9]:

BEHAVIOUR EXPLANATION - TELEO*: The piece of behaviour B has contributed to survival and reproduction because it was caused by a representation R such that, in the past, things of the same type caused a consumer subsystem to display behaviour B, and this contributed systematically to survival and reproduction.

This other version does not sound *quite as* (although, admittedly, it still sounds *rather*) thin. In any event, in the discussion to follow, I plan to grant Shea that teleosemantics can only provide BEHAVIOUR EXPLANATION - TELEO by way of explanation of a piece of behaviour[10].

Shea's solution is to add an extra necessary condition for a representation R to have content C: R must indicate that C[11], and its having done so in the past must figure in an explanation of the actual existence of R. In Shea's own terminology:

INFOTEL SEMANTICS: A representation of type R has content C if

IS1: Rs are intermediate in a system consisting of a producer and a consumer cooperating by means of a range of mediating representations (all specified non-intentionally), in which every representation in the range also satisfies IS1 to IS4;

9 We do not need to modify TELEO to provide for this reading: we just need to read the quantifier governing "a way" as taking wide scope.

10 Given that my main target is not Shea's argument against traditional teleosemantics, I will not discuss Millikan's rejoinder, in Millikan (2007), to the effect that, even if such explanations are thin, they are explanations nonetheless and, in fact, they are the kinds of explanations we are giving when attributing a content, or a function, to a device. I do agree with Millikan that explanations come in degrees, and there is no reason why BEHAVIOUR EXPLANATION - TELEO should not count in some contexts as a perfectly adequate answer to a request for information about a piece of behaviour. But, as I have urged throughout this and the previous chapter, and shall be presently stressing again, content attributions rely on an explanation of the existence of a mental state that is far more substantial than BEHAVIOUR EXPLANATION - TELEO -or Shea's alternative, for that matter. So, in my view, the point about the explanatory adequacy of traditional teleosemantics is moot.

11 Indication above only covers the case in which C is of the form *There is an F around S*. Here I am helping myself to the following notational variant from Indication: R indicates that there is an F around S iff R indicates instantiations of F around S.

IS2: Rs carry the correlational information that condition C obtains;

IS3: an evolutionary explanation of the current existence of the representing system adverts to Rs having carried information about C;

IS4: C is the evolutionary success condition, specific to Rs, of the behaviour of the consumer prompted by Rs. Shea (2007, p. 419 - I have renamed the four conditions)

The picture Shea has in mind, and its translation to the terminology I have been using, is the following:

- R is produced by some mechanism and consumed by some other (in a cognitive system, maybe, although not necessarily); R is one of several alternative representations that the mechanism produces. This is condition IS1. In the Democritus example we have been studying, R's producer is M, and R is, *e. g.*, M's being *on*, which is one of two alternative representations: M's being *on* and its being *off*.

- Condition IS2 (as Shea glosses it in his paper) amounts to $P(R|C) > P(R)$ in some local domain D, and this conditional probability being causally grounded. That is, if C is *there is an F*, and D is *around M's possessor*, condition IS2 amounts to M's being *on* indicating that there is an F around M's possessor.

- Conditions IS3 and IS4 amount to saying that the producer's Indication Profile (in IS3) and Fitness Matrix (in IS4), as seen from the perspective of the indication that C, are part of the evolutionary explanation of the existence of R.

That is, Shea's Infotel semantics may be regarded as, give or take, a notational variant of the BETTER DRETSKE theory of content I introduced in 1.2[12]. It is to be expected, then, that Infotel semantics displays the same kind of indeterminacy we found in BETTER DRETSKE + ETIOLOGICAL FUNCTION + INDICATION.

It does: take the following two attributions of content to M's going *on* in Democritus the frog: *There is frog food around* and *There is a black speck around*. Does infotel semantics warrants them?

- IS1: This condition is content-independent, and M's going *on* complies with it. As I said above, the alternative representation is M's going *off*. The producer is M itself, and the consumer is (say) Democritus's tongue-protracting device.

- IS2: *Ex-hypothesi*, $P(on|C) > P(on)$ for both contents.

- IS3: There is an evolutionary explanation of the current existence of M's *on* state -for instance, one along the lines of Etiological Theory- that adverts to its having carried the information that there was a black speck around M's possessor. On the other hand,

12 Shea (2007, p. 419, fn. 23) also remarks the congeniality of Infotel semantics with Dretske's theory of content. According to him, Dretske's theory of content relies on learning and not evolution, and is, therefore, not a version of teleosemantics. This is true for contentful states such as our beliefs and desires, but not so true for simpler contentful states such as those of some marine bacteria, frogs and the like (what Dretske calls *Type III Representational Systems*) in which content is indeed constrained by evolution, and not by learning, in approximately the way I described in 1.2.

there is another, equally adequate explanation that adverts to there having been frog food around M's possessor. Remember the gambit: both explanations converge in the same Fitness Contribution for M, and Fitness Contributions is all evolution cares about.

- I§4: Shea glosses this condition thus:

  > [A]n (historical) evolutionary explanation of the survival and reproduction of the representing system adverts to C's obtaining when Rs were tokened. Shea (2007, p. 419, fn. 22)

  Again, there are at least two evolutionary explanations available, etc.

2.3.2 *Etiosemantics and Behaviour Explanation*

In sum, Infotel semantics falls prey to the Indeterminacy Problem. This was probably to be expected: the proposal is engineered to comply with the explanatory-substantiality constraints that Shea wishes to enforce; but, as we have seen in earlier sections, whenever there is a substantial explanation of the existence of a representation of the kind Shea favours, many other, equally substantial explanations are also available.

This problem, it may seem, has no bearing on his project: after all, Shea is concerned only with the Behaviour-Explanation Objection, and his proposal does solve this problem: according to Infotel semantics, a state is not contentful unless it indicates that C, and this indication relation substantiates content-based explanations.

INDETERMINACY UNDERMINES EXPLANATORY RELEVANCE. But one could raise the issue that the Indeterminacy Problem seems to have consequences also in Shea's area of interest. Remember, he is in the business of defending that explanations such as BEHAVIOUR EXPLANATION are substantial:

BEHAVIOUR EXPLANATION The piece of behaviour B has contributed to survival and reproduction because it was caused by a representation with the content C.

If I am right and infotel content-attributions are multiply indetermined, the following worry appears to be in order: if two content-attributions C and C* to a representation R are warranted, in what sense does *its having content C in particular* helps explain behaviour B? It appears that content-based explanations abhor indeterminacy in the following sense: an explanation of behaviour B based on its having been caused by a representation with the content C must be, at the same time and for all alternative contents C*, an explanation based on its having been caused by a representation with the content C *as opposed to C**. Alternative content-attributions appear to undermine one another's claim to explanatory relevance in causing a piece of behaviour, in a way in which, *e. g.*, alternative function attributions do not: the answers "The alarm is ringing because you are smoking your pipe under a smoke detector" Millikan (2007, p. 440) and "The alarm is ringing because you are smoking your pipe under a fire detector" are not clearly incompatible.

THE EXPLANATORY RELEVANCE OF HPCS. Even if one wishes to defend that such undermining is, in the case of simple contentful states, benign at worst[13], etiosemantics remains a better package deal than infotel semantics: not only does the former not suffer from the Indeterminacy Problem while the latter does; moreover, it solves the Behaviour-Explanation Objection *better* than the latter: Shea added to teleosemantics the constraint that a representation only has the content that there is a fly around if it indicates that there is a fly around. The requirements for a representation according to Etiosemantics are stronger: it must carry correlational information, better or worse, about a bunch of property instantiations -those in the Cluster- and such that there is a mechanism that explains the recurrence of such properties in the environment of the representation's possessor. So, explanations that Infotel counts as substantial enough -e. g., those present in MANY EARTHS, MANY PEACHES- do not succeed in grounding a content-attribution according to etiosemantics. The latter needs more substantial explanatory facts to be available.

I suggest the conclusion of this discussion of Shea's theory, then, is that etiosemantics remains a better package deal. Not only does it solve an important problem (indeterminacy) infotel semantics does not solve, it has more stringent informational constraints so that, if infotel semantics solves the problem of the explanation of behaviour Shea was concerned with, etiosemantics, *a fortiori*, solves it as well[14].

2.4 PAPINEAU'S TELEOSEMANTICS

David Papineau (1987, 1993, 1998) has proposed a different approach to the indeterminacy problem. In this section, I first quickly reintroduce the Indeterminacy Problem using his terminology and -after providing a quick reminder of why, I think, my own account is not subject to the problem- I go on to present Papineau's solution, and raise two objections against it: the first is that, for his account to work, desires need to be individuated in a very non-standard way -*e. g.*, efferent nerves and maybe even hands need to be literally part of the desire for food-; the second is that the account only provides a content attribution for structures that have antecedently been identified as beliefs or desires. And it is unclear what exactly is to effect this identification: for example, in other accounts it may be suggested that beliefs are structures that share a particular kind of content, but this is not an option for Papineau -a certain state has some content or other partly in virtue of the fact that it is a belief. Without such an account of what beliefs and desires are, Papineau's proposal is incomplete.

2.4.1 *The Concertina Problem*

Papineau agrees with almost everyone in believing that there are multiple satisfactory explanations of the existence of certain biological traits:

> [I]magine that a species of highland antelope has some distinctive trait T which has been selected because it (a)

13 Shea does not feel that his project is constrained by common sense intuitions about content; this gives me reasons to think that his answer to the worry raised in the foregoing paragraph would be one of polite dismissal along these lines.

14 More on behaviour explanation in 3.10.

> alters the antelope's haemoglobin structure (b) increases oxygen uptake (c) enables the antelope to live on higher ground (d) gives it access to a plentiful food supply (e) increases reproductive success. Papineau (1998, p. 2)

This plurality of explanations, according to Papineau, gives rise to the *Concertina Problem* of function attribution: there is no fact of the matter as to which of these effects is T's function. Alternatively, we may understand this plurality as leading to an analogous function plurality: T is supposed to alter the antelope's haemoglobin structure, *and* increase oxygen uptake, *and*... In any event, this function plurality (indeterminacy), Papineau argues, leads to problems in consumer-teleosemantic accounts such as the following:

> The teleological theory [identifies] the content of an informational state with the circumstances in which it ... leads to advantageous effects. Papineau (1998, p. 3)

This proposal, according to Papineau, falls prey to the *Concertina Problem*: consumer semanticists, Millikan among them, propose that Democritus's M's being *on* represents *frog food*, "because it is only when food is present that the frog action has advantageous effects" -Papineau (1998, p. 3). But, Papineau continues, any other member of the Concertina of effects for M would do just as well: *health-preserver*, *reproduction-enhancer*, etc. The Indeterminacy Problem again.

Let me quickly remind why etiosemantics, if I am right, does not face a Concertina Problem: Certain instantiations of the properties *Being a health-preserver for frogs*, *Being a reproduction-enhancer for frogs* and the like are part of the property cluster that partly constitutes the HPC fly, just as instantiations of the property *Being a black speck* are. They, together with the homeostatic mechanism of fly-reproduction, will zero in on the very same natural kind. The concertina of effects is closely related to a concertina of properties that, in turn, can double as the seed of an HPC -see 1.4.5. This is the real kind to be used in the content attribution; thus, *There is a fly around*.

2.4.2 *Papineau's Solution*

Papineau proposes a different solution to the Indeterminacy (Concertina) Problem. Suppose we have a belief B, and we wish to provide a content attribution for it. Papineau believes we need to focus on one among the many "advantageous effects" appealed to in the quote above. One way to do so is to concentrate in the satisfaction conditions of the *desires* that the behaviour prompted by B helps to satisfy. We could render Papineau's proposal approximately as follows:

PAPINEAU: The content of a belief B is *There is a fly around* if its biological function is to be present in those circumstances in which the behaviour it prompts will satisfy the desire for catching a fly.

That is, Papineau proposes to modify the consumer-semantics insight according to which the content of a mental state are the Normal conditions for the state to help fulfil the function of its consumer, concentrating on one such particular consumer: the desire that the belief serves. Desires have a certain concrete content -involving, say, flies or food or black specks-, and we can profit from this determinacy in fixing the content of the beliefs that help satisfy them.

In earlier work Papineau considers some objections to this proposal, having to do, *e. g.*, with actions based on more than one belief, or false beliefs that nevertheless help satisfy desires -*cf.* (Papineau 1993, p. 74f). I propose to discuss here a more fundamental worry: this account will only work if there is a way of providing for determinate desire-contents, and this seems as difficult a task as providing for determinate belief-contents. Are we back to square one, then?

Of course, Papineau is alive to this worry. His strategy to provide desires with determinate contents is an adaptation of Neander's low church teleosemantics (see 1.3.2). We could render his proposal regarding the content of desires as follows:

LOW CHURCH DESIRES: A mechanism M's positives are a desire for Fs if, at the lowest level of functional analysis at which M is an unanalysed whole, it is supposed to bring about the acquisition of Fs.

The idea is to think of a cognitive system as a set of nested boxes. A particular desire is a box which contains other boxes, and which sits inside others[15]. The content of a desire has to do with the function of the desire's box, independently of the other boxes it has inside (thus the *unanalysed whole* restriction), or the box/es inside which it sits (thus the *lowest level of functional analysis* restriction).

So, consider a desire of which we want to say that it's a desire *for food*. Such a desire (which we may assume is a certain physical state) has a concertina of selected effects (Papineau 1998, p. 11): that you move your arm, that the spoon enters your mouth, that you acquire food, that the food be digested etc. LOW CHURCH DESIRES should single out *that you acquire food* from the rest of effects. According to Papineau, it does:

> The initial stages in the concertina of results [*e .g.*, that you move your arm, that the spoon enters your mouth] depend as much on the beliefs behind your behaviour as on the desire itself. (If you didn't believe that there is food in your spoon, your desire for food wouldn't make you put the spoon in your mouth...) Papineau (1998, p. 12)

Papineau suggests that we deal with this part of the concertina by asking that the content of the desire be one among the effects that the desire is *always* supposed to produce -not just in combination with this or that belief. Now, to deal with the rest of effects in the concertina (*e. g.*, that food be digested, that you survive and reproduce):

> If we apply Neander's analysis, then we see that the function of acquiring food is nevertheless specific to this desire. For the non-fulfilment of the further functions, like digestion and reproduction, doesn't show that this desire is malfunctioning, since these further functions depend not just on the desire doing its job, but also on other traits, like the digestive and reproductive systems, doing their jobs too. Papineau (1998, p. 12)

That is, the effects to be screened off are, in fact, effects of a larger system that includes the desire together with, e. g., the digestive and

15 The analogy gets awkward when one is forced to think about a box sitting inside *two* different boxes such that one is not contained in the other, but I hope the idea is clear.

reproductive systems. After ruling these out, *acquiring food* is the only effect that remains a candidate to be used as content for the desire, as we wanted.

Mendola (2006) has claimed that this account of the content of beliefs and desires is circular: the strategy for screening off part of the undesired effects makes an appeal to "the beliefs behind behaviour" (see above) and, particularly, to the content of these beliefs: "if you didn't believe *that there is food in your spoon*, your desire for food wouldn't make you put the spoon in your mouth" (my emphasis). But the content of beliefs was supposed to depend on the content of desires; if the content of desires depends on that of beliefs, we launch a vicious circle. Or, at least, as Mendola (2006, p. 314) points out, a "spiral into the past" where the content of beliefs depends on that of desires, which depends on that of previous beliefs which depends...

It is difficult to see what Papineau's rejoinder could be. Some of his writings -I'm thinking of (Papineau 1993, p. 73)- suggest that he may want to propose that his analysis of the content of beliefs and desires should be applied *simultaneously* to every belief and desire in the spiral, in a diachronic ramsification of sorts. Much more would need to be said about how to make this work, though. In any event, I have a different worry to press: if Papineau's proposal is to stand, we need to *individuate* desires in a highly non-standard way.

2.4.3 *Individuating Beliefs and Desires*

Consider again the reason Papineau provides for leaving out of a desire's content effects such as *food being digested*: food may fail to be digested and the desire still be functioning correctly, because the failure may be happening elsewhere; in the digestive system, say. If this is the right way to keep food being digested out from the content of desires, it must be that the analogous maneuver is not available for the effect *acquiring food*: if the desire is tokened and food fails to be acquired it must be because the desire has malfunctioned.

For this to be the case, the desire itself must stretch out all the way to the motor control areas, the efferent nerves and beyond, up to the hand. Otherwise, one could deny that acquiring food is to be counted as the content of desires in the same way that digesting food was filtered out. Paraphrasing Papineau, one could say that

> the non-fulfilment of the function of acquiring food doesn't show that this desire is malfunctioning, since this function depends not just on the desire doing its job, but also on other traits, like the efferent nerves and the hand muscles, doing their jobs too.

That's why I was suggesting that desires appear to be strange beasts in Papineau's philosophy of mind: they quite literally stretch down my arm to the tip of my fingers; and if I suffer demyelinisation in the efferent nerves connecting my brain to my right arm, my desire for food is *physically* changed[16].

16 In fact, I think, Neander would not agree with Papineau that the specific function of a desire may be something like *acquiring food*. If we are to judge from what she says about the frog's M mechanism, it is likely that she would take the function of a certain desire to be some of its effects in neighbouring brain mechanisms; acquiring food seems, indeed, to need of the concerted effort of a great part of the human body, well beyond a certain mental state or other.

This is an implausible consequence, and an important drawback to his theory as it stands; but it also points to a more general shortcoming of his account of content, which has to do as well with the individuation of beliefs and desires. Papineau's account aims at providing a content attribution to a certain structure *that has been independently characterised as a belief*. The account does not comment on what makes a certain brain structure a belief, or a desire. Rather, it takes as a given that it is one or the other or neither. But, if a theory of content of the kind Papineau is interested in developing does not establish this kind of facts, it is difficult to see what will. For example, in a theory such as Millikan's, one can characterise beliefs as the mental states that have a content of a certain kind -maybe with some other structural constraints- and do the same with desires. This is not an open option for Papineau because, according to his theory, something does not have a determinate content or other unless it is already a belief, or a desire.

Papineau cannot easily help himself to the traditional functionalist idea according to which something is a belief if it has a causal role of a certain kind. The combination of this view with a teleosemantic account of content is unstable: for example, it would turn the swampman -a physical replica of Donald Davidson that has emerged in a swamp as a result of the purely random recombination of molecules- into a creature with as many beliefs and desires as we have, but such that all of its beliefs and desires are empty -not just gappy, mind you, but *blank* through and through. That would be a monster generated by theory if anything is[17].

Papineau has not provided a theory of the content of beliefs until he has provided a theory of what makes a certain structure a belief -and it is unclear how he could go about providing that. As it stands, his theory does not seem to deal satisfactorily with the Indeterminacy Problem.

2.5 MORE ON HPCS

Once a number of basic tenets of the theory have been presented, and some of its differences to some contemporary teleosemantic approaches charted, I wish to go back to the characterisation of HPCs. In the present section I introduce what we may call disjunctive HPCs, and the possibility of reference change. I also discuss some of the relations between traditional-essence real kinds and HPCs as I have defined them. Finally, I draw a clearer distinction between the property *Being an F*, where F is an HPC, and traditional properties such as *Being a red square*.

2.5.1 *An Explanation for Several Semantic Phenomena*

DISJUNCTIVE CONTENTS In ARETE AND THE PEACH TREE I stipulated that Ma's abduction is a one-off situation in the story of Arete's

17 ... Although I cannot help but note the Beckettian beauty of such a swampman. For Papineau's take on the swampman see Papineau (2001). In this paper he defends that we need not worry about swampmen as far as they remain safely confined in counterfactual situations. Míguel Ángel Sebastián has suggested to me an elaborate way in which something sufficiently similar to a swampman may actually exist -involving a DNA sequencer hooked up to a random-number generator, and advanced *in vitro* technology. It would be interesting to explore the import of such examples for Papineau's views on the swampman. This will need to be matter for future work, though.

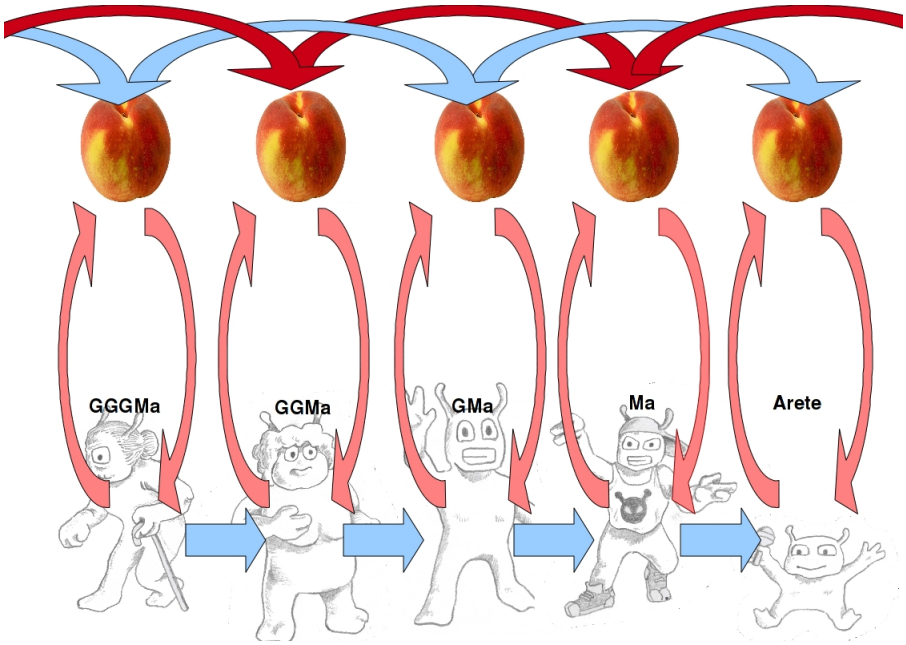

Figure 4: Disjunctive Contents

family. Consider, instead, a scenario in which roughly half of Arete's ancestors had fed on cheaps (*cf.* fig. 4). What is the content of M's being *on*, then?

In this situation we have a cluster of properties that is held together by two different specialised homeostatic mechanism. One of them, together with the property cluster, individuates the real kind *peach*; the other, the real kind *cheap*. In such a case we probably wish to say that the content of M's being *on* is, as the content of JADE is, *disjunctive*: *There is a peach-or-cheap around.* A simple way to accomplish this is by introducing the following notation

DISJUNCTIVE HPC: Given a set of properties P, if there are i sets of specialised homeostatic mechanisms SHM_i that explain the fact that, in a domain d, the properties in P are frequently coinstantiated, then there also is a *disjunctive HPC* individuated by the set of every HPC F_i which, in their turn, are individuated by:

- The smallest set of properties P_i' such that SHM_i explains the fact that, during a certain time period t and in a certain spatial area a, the properties in P_i' are frequently coinstantiated, and
- SHM_i.

To ensure that disjunctive HPCs do what we want them to do, we have to tweak TFA3 in THERE IS AN F AROUND, to rule that a state has a content involving disjunctive HPCs only if

1. Every SHM_i explains to a certain, sufficiently$_1$ high degree, the fact that the properties in P are frequently coinstantiated throughout the generations needed for selection for M.

2. Only the whole set of SHMs explains to a sufficiently$_2$ high degree, the fact that the properties in P are frequently coinstantiated throughout the generations needed for selection for M.

Where sufficiently$_1$ marks the point at which we want to say that a certain HPC has a non-negligible participation in the explanation of the existence of M, and sufficiently$_2$ marks the point where we want to say that the existence of M is satisfactorily explained. There are paradigmatic cases of reliance in a disjunctive HPC, the situation depicted in fig. 4 being one of them. In that case we may say that M's being *on* has a content involving the disjunctive HPC *peach-or-cheap*. There are also paradigmatic cases in which there is no reliance in an HPC. Take, again, MANY EARTHS, MANY PEACHES. Here the notation just introduced allows us to talk of the disjunctive HPC *peach*$_1$*-or-peach*$_2$*-or-...-peach*$_n$. But such disjunctive HPC is not to be considered as part of the content of M's being *on* in the story: *no SHM*$_i$ explains to a certain, sufficiently$_1$ high degree, the fact that the properties in *P* are frequently coinstantiated throughout the generations needed for selection for M.

Of course, it may be that the properties of *Being sufficiently*$_i$ *explanatory* are vague properties and that, for some mental states, it is unclear whether a certain disjunctive HPC is or is not part of its content; or whether the mental state has content at all.

REFERENCE CHANGE. If, after a sufficient number of generations of feeding on peaches, Arete's family starts feeding on cheaps and keeps doing that for ever more, after some time we may wish to conclude that M's being *on* has stopped meaning *There is a peach around* and started meaning *There is a cheap around*: given that peaches are nowhere to be found in Arete's environment, genetic drift would have ended up mutating away from M; but M sticks with Arete's family because it is still useful. Only the explanation of its usefulness has changed. Now it must make reference to a different real kind.

This would not be a well-described possibility if a mental state was individuated by its content. Luckily, they are not, or shouldn't be. Plausibly, the right way to individuate mental states is what Millikan (1984, p. 23f) calls *reproductively established families*. Without entering in the painstakingly detailed treatment of the notion in the *op. cit.*, the idea is that what makes a certain mental mechanism a token of the type M is that it derives from other M tokens through a process similar to reproduction[18] that preserves a certain reproductive character, that is, a relevant set of properties, which may include M's causal profile -what causes it to go *on*, and what its going *on* causes. Belonging to a certain reproductively established family is what makes a certain mental mechanism a token of M.

The independence of M-hood from content, together with the fact that ETIOLOGICAL FUNCTION is a modern-history theory of function (cf. the discussion of condition EF1 in 1.2.3) is enough to accomodate the possibility of reference change.

2.5.2 *Traditional-Essence Kinds vs. HPCs*

It is useful to distinguish two types of real kinds. On the one hand, those such as *water* or *gold* that have what may be called *traditional essence*. In such real kinds a *hidden essence* accounts for all or most of the visible

18 It is not the commonsense notion of reproduction because, e. g., it has to make room for the fact that a legged daughter may be born from a mother without legs. For further details see Millikan (1984).

properties which are normally associated with them. Thus, the property *Being made of H_2O molecules* helps account for the transparency, liquidity at ambient temperature, disposition to dissolve ionic compounds, etc. of water.

On the other hand, there are kinds such as *tiger* or *mammal*, used in biology, and maybe even others such as *Renault Megane -cf.* Millikan (2000)- or *good -cf.* Boyd (1988). The problem of the essence of such natural kinds is perceived as much more pressing; to the point that it's natural to conclude that they have no hidden essence at all. The insight I'm adhering to by using the notion of HPC is that a causal mechanism that keeps properties together may be enough to bring a real kind into existence, even in the absence of a traditional essence -although, as I said in 1.4.5, traditional essences may be an essential ingredient of some HPCs.

It is unclear that creatures as simple as Democritus or Arete, according to my account, can secure reference to traditional-essence real kinds such as *water*. This is so because something's being water -that is, being H_2O- cannot fulfill the whole explanatory bill etiosemantic demands for content attributions to be justified. Being water, in the traditional sense, fails to explain the sufficient density of instantiations of properties in the cluster (transparence, liquidity, etc.) in the domain in which a certain mental state may be selected for. But, without an explanation of such density -*i. e.*, of the frequent co-ocurrence of properties in the cluster- we do not have enough grounds to attribute content in the simple cases we have been seeing[19].

The closest analogue to water that a creature such as Democritus can entertain is an HPC in which the traditional essence *Being H_2O* is supplemented with a mechanism that explains that molecules of H_2O keep appearing around Democritus in the spatiotemporal region in which an H_2O indicator is selected for. In the case of water it may be something similar to the water cycle: water recurs in the environment of the agent -a nearby river does not run dry, for instance- because water downstream is heated and travels upstream as vapour. A story should be told -rather, some story should be the true cause- why water keeps happening around the agent. So, the real kind closest to water that one of these simple creatures will be able to entertain contents about is an HPC whose specialised homeostatic mechanism is constituted by being H_2O together with the water cycle. Maybe a not totally unfitting English name for such an HPC is *Earth water*.

A similar thing would happen with gold. Imagine a population of microorganisms that use gold for some purpose -there is none, that I know. If an indicator with *Being shiny and yellow* as input mutates into existence in one of these creatures it may get selected for, and we may want to say that the indicator's being *on* means *There is gold around*. For very simple creatures, this will not be the right content, but, rather, *There is gold* around*, where gold* is individuated by a homeostatic mechanism constituted by *Having atomic number 79* and whatever it is that explains that more gold is sufficiently near most amounts of gold in the environment of these microorganisms. Maybe a name for such an HPC would be *This chunk of gold*.

19 Beyond a certain threshold of sophistication, content-crunching systems, such as the ones introduced in chapter 4, will not have this particular limitation because some of their contents will depend on homeostatic mechanisms, such as physical laws, which are universal.

Fly and *hoverfly* are special in that they are HPCs which enjoy an English name; and it may not be a coincidence that the most common examples of very simple contentful states discussed in the literature on content are about these kinds of entities. The lesson to draw from this discussion is that reference to traditional-essence kinds is a more sophisticated cognitive feat than reference to HPCs. I think, in fact, that the phylogenesis of contents is perfectly upside down from what Russell-style empiricists would have hipothesised: the first contents to appear involve HPCs; then come contents involving traditional-essence real kinds; and only then contents involving properties such as *Being red* or *Being square*. I will be saying nothing about the emergence of contents involving the latter properties in this dissertation. That will remain matter for future work.

2.5.3 *Kinds and Properties.*

The thing to which the most basic contentful states refer, I have been arguing, are HPC real kinds. States with similar intrinsic complexity as the ones I am crediting with content -*e. g.*, states that have the same causal powers under some suitable description of their intrinsic behaviour, such as INPUT in 1.4.1- are contentless precisely because they do not owe their sustained existence to a suitable relation with a real kind. This is, for example, the case with M in MANY EARTHS, MANY PEACHES.

Let me call *Shoemakerian properties* those individuated by their intrinsic causal contributions, along the lines of Shoemaker (1998):

> Any property has two sorts of causal features: "forward-looking" ones, having to do with what its instantiation can contribute to causing, and "backward-looking" ones, having to do with how its instantiation can be caused. Such features of a property are essential to it, and properties sharing all of their causal features are identical. Shoemaker (1998, p. 59)

I am unsure that *Being an F*, where *F* is an HPC, is a Shoemakerian property: instantiations of *Being an F* can only be caused by the specialised homeostatic mechanism in play, from other instances of the very same property -or suitable ancestors thereof, up the phylogenetic stream. So, the backward-looking features of an instantiation of *Being F* are episodes of further instantiations of the same property causing it, via the homeostatic principle. We can formulate in a compact fashion said backward-looking features, together with its forward-looking properties. Suppose that the HPC F is individuated by a set of Shoemakerian properties *P'* and a specialised homeostatic mechanism *SHM*:

SHOEMAKERIAN VERSION OF BEING AN HPC: λx[an instantiation of *x* caused its instantiation, via the specialised homeostatic mechanism SHM; and it causes other instantiations of *x* and events of type *t*]

Where events of type t are whatever events the instantiation of *Being an F* normally causes -prey of F to abandon the scene, for example, or predators to appear.

This Shoemakerian version does not fix what it is to be an F. This is because the HPC *F* is a concrete, spatio-temporally situated entity,

not completely defined by its causal profile: it is the sum of the particular instantiations of a cluster of properties -these could very well be Shoemakerian- together with the homeostatic mechanism that links them. Other things, replicated in the same way from the same homeostatic mechanism, will not be *that very* HPC.

This may lead to conjecture that the most basic mark of a contentful state may be that the explanation of the existence of the state involves, in the ways I have been detailing through the chapter, a concrete HPC, and not just a collection of Shoemakerian properties. And we may, subsequently, reformulate our solution to the Indeterminacy Problem. There is a clear sense in which all that frogs care about are the forward-looking causal features of *Being a fly*: these are the features that account for frogs detecting flies, and for frogs being fed by flies. And Fodor is also right that a number of real (and not so real) kinds are such that the property of *Being a fly* has the same forward-looking properties as the property of being those other kinds has, in the environment of the agent (e.g., *Being a black speck* and what have you.) What Fodor and other fans of the Indeterminacy Problem failed to see is that the fact that such forward-looking properties recur in the environment of the agent is an indispensable part of the explanation of the persistence of the contenful state, and this recurrence is explained by an homeostatic mechanism that, together with the property instances Fodor recognizes, fixes an HPC once and for all. No indeterminacy remains, because if *Being an F* -where F is an HPC- is the property that fits this explanatory bill, there is no alternative property *Being an F** -where F* is an HPC- that does, too.

2.6 RELATED OBJECTIONS

Before finishing this chapter, and with it this first, long exposition of the very basics of etiosemantics, I wish to show how the theory deals with a couple of prominent objections to teleosemantics.

2.6.1 *Objections to the* Real-Kinds *solution*

Etiosemantics provides grounds for one attractive idea explored in the literature on teleosemantics: the suggestion that the explanation that is to ground content attributions must appeal to the detection of real kinds -*cf.* Sterelny (1990). We don't need to appeal to a, seemingly, *ad-hoc* preferability of real kinds as intentional objects for simple mental states. Rather, it is *only* real kinds, and not other, less natural entities, that enable the selection for indicators. This is why it is real kinds that must figure in content attributions to basic mental states.

Nevertheless, doubts have been raised about the suggestion that the most basic contents involve real kinds: according to Neander (1995), naturalistic accounts of content ought to account for the existence of basic contents that involve non-real kinds:

> Birds have an innate preference for ripe fuit, and sometimes for mates with vivid red tails, so presumably they have representations of colors, and so color kinds had better count as natural kinds, or we will have shaved too much. But in virtue of *what* would they count as natural kinds? (*op. cit.*, p. 127)

First of all, we should point out that there is nothing wrong with the idea that colours are real kinds. There may actually be homeostatic mechanisms that explain that colours recur in our environment. Of course, these homeostatic mechanisms will not be genetic as with flies, but, as in the homeostatic mechanisms that individuate clouds, have to do with the recurrence of colour-forming conditions: physical laws explain that surfaces recur in our environment, and that these surfaces have reflectance properties. Colours would then be identified with such reflectance properties. This kind of theories[20] are surely controversial, but they are not obviously wrong.

More importantly for our current purposes, it is not obvious that the birds in Neander's example have colour-involving contents. Neander's argument could be reconstructed with the help of a concrete example: Protagoras, a bird, has a mental mechanism M. M does the following: whenever there is thus-and-so kinds of light, Protagoras goes towards it. Since the mutation took place first in Protagoras's family, it has helped Protagoras's ancestors to thrive by either dragging them towards ripe fruit or to suitable mates. What is the content of M's being *on*, then? There is nothing in common between suitable mating partners and ripe fruit but their displaying vivid colours, so it must be this that M detects. Now, *being a vivid colour* is not a natural kind, etc.

But, despite Neander, the content of M's being *on* is not whatever ripe fruit and mating partners have in common but whatever it is that explains the evolutionary value of M for Pierre and his family. The properties that causally ground M's Fitness Contribution are *Being a vivid colour*; *Having such-and-such nutritional properties*, *Being suitable for mating*, etc.. There is no homeostatic mechanism that links all of these properties together, so this is one of the cases in which it makes sense to make M's being *on* a state with content that involves a disjunctive HPC (see 2.5.1): *There is a ripe-fuit-or-suitable-mate around*. That is, if there really is a single mental structure M that channels responses to ripe fruits and suitable mates and nothing else.

Contents involving such gerrymandered disjunctive HPC such as *ripe-fruit-or-suitable-mate* are, probably, comparatively infrequent: we need a case in which two or more homeostatic mechanisms are present and active in the evolution of some creature. Then again, if they are not infrequent, then they aren't. The world, and not philosophers, is to decide.

2.6.2 *Pietroski's* Kimu

In Pietroski (1992)'s famous example, the *kimu* are colour-blind creatures until Jack, a mutated baby *kimu*, develops a certain mental structure, B. B works such as this: whenever there is red nearby, Jack feels compelled to go towards it. Because of this, he is dragged to the top of a nearby hill every morning, to spot the lovely reds of the rising sun. The *snorf* are predators of the *kimu*, and normally go hunting at the time that Jack is uphill. This has saved him from being caught, and, thus, has allowed him to reproduce. This is the only explanation why Jack's sensitivity to red gets to be selected by Nature.

The intuitively correct content to be attributed to B is *red*. Or, at any rate, according to Pietroski, it is obvious that "B-tokens are not about snorfs" (Pietroski, *op. cit.*) Nevertheless, consumer semantics, looking at

20 Tye has developed one such account in his Tye (2000).

whatever it is that explains the correct performance of the consumers of representation B, would probably want B to have the content *snorf-free area*, or something to that avail. Indeed, Millikan (2009) suggests that the content of B is *fewer snorfs this way*.

One of the reasons why we may feel uneasy in ascribing content to B is that the homeostatic mechanism at play is extremely fragile: there is causally-grounded correlation between red and fewer snorfs only in one place at one time of the day. Besides, this brittle correlation must remain operative for a sufficient sizeable number of generations in Jack's family line, if there is to be selection for B. This suggests that kimus have a not very flexible behavioural repertoire, are fairly basic creatures and, *contra* Pietroski, we shouldn't feel too inclined to grace them with contents involving the property *Being red*.

In any event, etiosemantics provides a reasonable content attribution, somewhat more intuitive than Millikan's -although our intuitions about this case are feeble, and Millikan's candidate seems OK if intuition grounds are alone to be considered- but more realistic than Pietroski's:

The properties co-ocurrence of which help ground the Indication Profile and Fitness Matrix of B's producer include, among others, *Being closer to red* and *Being snorf-free* (or, better, *Having less snorfs that areas in which other competing kimus are*). Is there any homeostatic mechanism that links properties in this cluster together? There is; it's something like this: there is a place close to the habitat of kimus that is higher than the rest of the nearby territory. Its being elevated grounds the correlation between *Being closer to red* and *Being comparatively snorf-free* (at dawn). It remains a causal ground for the correlation during the time needed for selection for B because orography tends to stay put. The HPC that, according to etiosemantics, should figure in the right content attribution to B is individuated by:

- All the instantiations of properties that Being on top of the hill have caused: *Being snorf-free*, *Being closer to red than in the valley*; but also, *Being colder than in the valley*, *Being rocky*, etc.
- The homeostatic mechanism constituted by: the fact that the hill is higher than the valley and the fact that orography tends to stay put.

It is not clear that this HPC has a satisfactory name in English. The best we can do is, probably, *The top of this hill*. When Jack and his family are affected by red light, they go towards it, because that is, by their hopelessly dim lights, like going to this hilltop. Most of the times, they are wrong but, crucially, they are right at a time in which snorfs abound in the valley. Hence B's fitness contribution.

Pietroski is right in refusing to equate too directly the content of a mental state with whatever it is that makes it evolutionarily useful. I'm advocating that the right content attribution involves the HPC that explains the recurrence of the conditions that make it evolutionarily useful. The ascription of content we get in this way is maybe more plausible than Millikan's "fewer snorfs this way". Agreedly, it is still superficially less plausible that a simple ascription of *red*, but the plausibility of this latter attribution seems to depend on the anthropomorfic fallacy of ascribing too quickly a counterpart of our red-involving contents to kimus.

3

EPHEMERAL CONTENTFUL STATES

In the first two chapters I have introduced the basics of a naturalistic theory of content I have called (with a respectful nod to *teleosemantics* and *biosemantics*) *etiosemantics*. Up until now I have been dealing with extremely simple mental mechanisms. One of such mechanisms, M, can be in one of only two possible states: *on* and *off*. The class of properties such that M responds to their instantiations as a matter of nomological necessity we have been calling its *input*. M is also hooked up with some other systems, so that its going *on* causes a number of changes in them. This is M's *output*.

M has been selected because its input and output have conspired to make it fitness-contributing for its possessor throughout generations -*cf*. 1.2.3. There is an explanation of this fact, which has to do with the sufficiently frequent coinstantiation of a number of properties around M's possessor, and the causal processes that have ensured this frequent coinstantiation during the time needed for selection for M -*cf*. 1.4. The etiosemantic proposal is to attribute M's being *on* with a content that involves the natural structure individuated by those property instantiations and these causal processes -an HPC (*cf*. 1.4.5). In this chapter I will start developing an account of the content of mental states that have not been selected for -or *ephemeral states*, as I will also call them. For example: Democritus's [M's being *on*] mental state has happened a sufficient number of times for its own selective history to play a content-fixing role. The states I will start discussing in this chapter -and further discuss in the next- do not have this feature.

To begin with, in section 3.1 I give a content-attibution recipe for what we could call (not ephemeral, but) selected *proto-beliefs*, in the same sense that M's being on could be called a proto-judgement (the suitability of these labels is briefly discussed in 3.1.2). These proto-beliefs (whose implementation involves what, after Dretske, I will call *property recruitment*) will be about the fact that certain two HPCs are causally related in a certain way.

In 3.2 I start developing the account that allows the attributions of contents to mental states that are a result of a property recruitment that has *not* been selected for. That is, the content attribution I provided in 3.1 for selected-for states, I provide here for ephemeral ones. In this connection, I discuss Millikan's theory of derived proper functions, also designed to account for ephemeral states. I express some misgivings about the theory, and canvass the way in which I propose to fix them, for the purposes of a theory of content. After that, in 3.4, I provide a concrete example: a simplified version of the mechanism of *long term potentiation* in our brains.

After that, I discuss contents involving individuals. I follow the same sequence as in the first part of the chapter: first (3.7) I offer a recipe for the attribution of individual-involving contents to selected-for states; after that (3.8) I extend the idea to ephemeral states; finally (3.9) I discuss a concrete example. After the first part of the chapter (about contents involving causal relations) and the second (about individuals) I discuss several normative notions -Procedures (3.6) and Norms (3.10)-

which should help alleviate anxieties about the role of contentless states in the explanation of behaviour and, in general, should help account for our internalist intuitions.

One of the main results of this chapter is that it is possible to attribute content to ephemeral states without attributing function to them. This is interesting, and maybe desirable, given the increasing suspicion with which pan-adaptationist accounts of mentality (according to which all mental states are adaptations) are regarded. While it is unclear whether Millikan's account of derived functions implies that such functions are adaptations, it is informative to see how one can go about providing content attributions for functionless items, while still following the main teleosemantic insights.

3.1 PROPERTY RECRUITMENT

Discussions of the Homeostatic Property Cluster theory in the philosophy of biology have normally stressed those members of the Property Cluster that scientists have traditionally relied upon in order to identify a real kind. Say, the properties of zebrahood or tigerhood that first come to our mind -stereotypical properties such as *Having black and yellow stripes*, which, in less enlightened times in the philosophy of biology, were taken as criterial for the presence of some real kind or other. Our entry point to HPCs, instead, has been the (proto-)epistemic relations of very basic cognisers such as Democritus or Arete with their environment. From this perspective, instantiations of properties in the Cluster are naturally understood as closely related to what more sophisticated creatures will take as *evidence* of the presence of the kind.

Now, the presence of a certain real kind may be a very good sign of the presence of another. Consider the case of a real kind F that, in a certain domain, is instantiated whenever another real kind G is instantiated. This may be so because the presence of F be sufficient *tout court* (*i. e.*, metaphysically sufficient) for the presence of G. Alternatively F may be just *in situ* sufficient for G: the causal underpinnings of the relation between F and G may be available only in a very restricted context, place or time, *e. g.*, only around the agent's neighbourhood, or only during weekend evenings. We may say that

SUFFICIENT: F is sufficient for G iff

1. $P(G|F) \approx 1$ and
2. The probability in 1. is causally grounded.

Suppose M is a mechanism such that THERE IS AN F AROUND warrants an attribution to M's being *on* of the content *There is a G around* -I will also call such mechanisms *G-mechanisms*. Suppose also that M has some property E as its input. We will say that M has *disjunctively recruited* the real kind F if it changes its input to $(E \vee F)$ (*cf.* figure 5)

If the instantiation of F is (*in situ*) sufficient for the instantiation of G, then disjunctively recruiting F will improve M's indication profile as seen from the perspective of the indication of Gs -on condition that E is not necessary for G, and F is not always uninstantiated[1]. That is, *disjunctively recruiting* a sufficient real kind is, normally, rewarding.

1 In fact, we don't need P(There is an F around|There is a G around) to be exactly equal to 1 for the recruitment of F in the input of a G-mechanism to be rewarding. In the Appendix C I give rigorous conditions for the recruitment of a property to be rewarding.

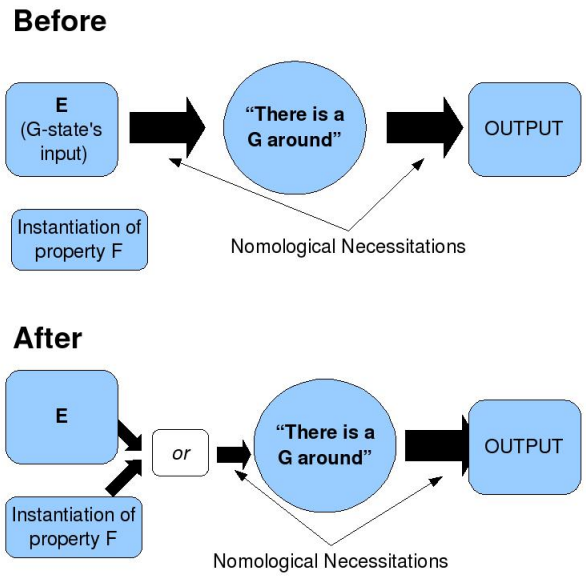

Figure 5: Disjunctive Property Recruitment

3.1.1 *Recruiting a Nonnomic Property*

Property recruitment, as just defined, is a process at the end of which some real kind enters in a nomological relation with a mental mechanism: the instantiation of F, among other things, causes the mechanism to fire, as a matter of nomological necessity. If this is so, recruiting a property for the input of a state is only possible in the cases in which such properties are, as Fodor (1986) has called them, *nomic*: properties that can enter directly in nomic relations[2]. According to Fodor there are many other, *nonnomic* properties, which can only enter in nomic relations indirectly, through nomic properties. For instance, *Being a crumpled shirt* (Fodor's example). The property of being a crumpled shirt that my shirt has can cause me to go iron it, but not directly; only through, say, geometrical and colour properties of the shirt.

However plausible this is for the case in which *I* am caused to iron something[3], it *is* quite plausible that simple mental mechanisms only enter in causal relations with properties such as *Being a crumpled shirt* through other, more proximal properties. If this is true, it means that mental mechanisms cannot simply recruit nonnomic properties in the crude way we have discussed; they have to do it through the recruitment of mediating nomic properties.

More interestingly for our purposes, it may be argued that the property of *Being homeostatic property cluster Q* is nonnomic in Fodor's sense:

2 It may be that there are no nomic relations at all between mental mechanisms' going *on* and other properties: on this view, the firing of a mental state would need a very complicated set of conditions that cannot be described without heavy use of ceteris paribus clauses. We may go around this issue by saying that, of all the properties that a mental state indicates, there is one property, Q, such that $\forall i P(on|Q) > P(on|Pi)$, the posterior probabilities being grounded, in Dretske's sense. That property has a psychophysical relation with M that is close enough to being nomological, for our purposes.

3 Not very plausible, I think, but the metaphysics of causing *people* to do something are, to be sure, extremely complicated.

whatever an HPC causes at t, it is because some of the Shoemakerian properties in its cluster causes it at t, the historical properties that tie the cluster together not being causally efficacious at t. This amounts to saying that HPCs only enter in indirect nomic relations: they are nonnomic[4].

In the process of recruiting, a mental mechanism can go around the nonnomicality of a property in different ways. The simplest way is to recruit a nomic property that indicates the nonnomic one that is ultimately interesting. This is just the case with *Being a fly* and *Being a moving black speck*: *Being a fly*, we hypothesise, cannot directly cause Democritus's M to go *on*; and Nature goes around this issue by placing a nomic property that is a good indicator of flyhood (blackspeckness) as M's input. This way of going around nonnomicality lies at the centre of the etiosemantic account of mental content, and of other broadly teleosemantic proposals. I wish to suggest now that one special case of this way of recruiting nonnomic properties is a plausible building block of contents involving causal relations among HPCs:

In the cases in which F is a reliable sign of G-hood, the positives of a mental mechanism which have content *There is an F around* (I will call such a mechanism an *F-mechanism*), if they are available, will normally also be a good indicator of G-hood: not in vain, what explains the existence of an F-mechanism is, in part, that it is a good indicator of the presence of F[5].

A causal link between the F-mechanism and the G-mechanism such that whenever the former fires the latter does too (*i. e.*, disjunctively recruiting F-mechanism's firings for the G-mechanism's input) may, therefore, be useful: it means that the G-mechanism has recruited a new, quite reliable indicator of G-hood. Alternatively, it may be understood as meaning that the survival value of the F-mechanism has been enhanced -now, as part of its output, it also helps to bring about whatever benefits accrue from the successful detection of G. See figure 6.

If there is a way to replicate this causal link accross generations -*e. g.*, if the recruitment is the result of a heritable genetic mutation- then the extra fitness contribution may end up fixating the link in the population of its possessors. This explanation of the existence of the mental state jointly formed by the F-mechanism, the G-mechanism and the causal link among them may ground a content attribution for it. Let us call this state FG.

It is a natural extension to the etiosemantic main insight to defend that FG will be contentful if there is a (higher order) HPC, say, Q, that has in its Cluster the instantiations of F-hood and G-hood that have made FG useful for its possessor and its ancestors, and explains that $P(G|F)$ has remained sufficiently*[6] high across the generations needed for selection for FG. In this case, FG's content will be something like *The causal mechanism SHM ensures that when there's an F around there tends to be a G around*. This is so because Q (the higher order HPC) is constituted,

4 *Cf.* section 2.5 for a fuller elaboration of these remarks.

5 Of course, as I already pointed out in 2.2.2, *Being a good enough indicator of* is not a transitive relation, and it may well be that the F-mechanism is a good indicator of F-hood, F-hood a good indicator of G-hood, and the F-mechanism not a good indicator of G-hood. The discussion to follow is relevant only for those (many) cases in which the relation of good-enough indication does carry over. In Appendix B I give the formal condition for such a recruitment to be rewarding.

6 I use *sufficient** instead of *sufficient* to remind us that there is a rigorous understanding of sufficiency available, in terms of Fitness Contributions, and their components -Indication Profiles and Fitness Matrices.

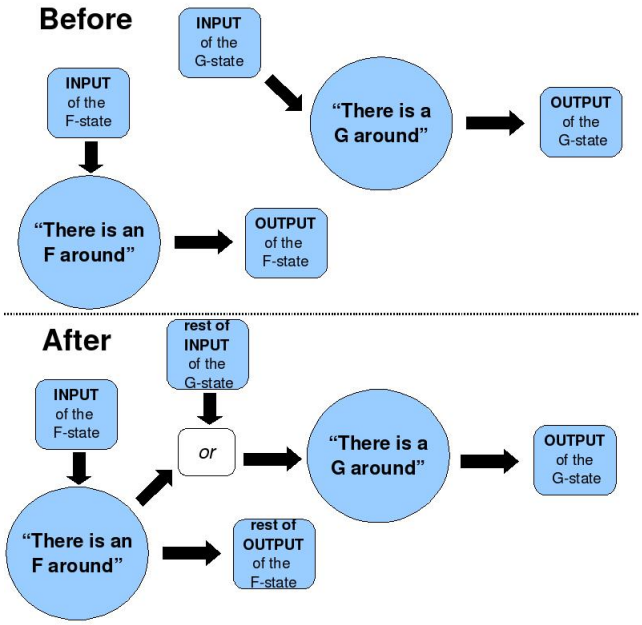

Figure 6: Disjunctive Recruitment of an F-mechanism

inter alia, by the instantiations of F-hood and G-hood around S that were kept together by a certain specialised homeostatic mechanism SHM that keeps the conditional probabilities as needed. More formally:

SHM BETWEEN FS AND GS: A subject S has a mental state FG with the content *The causal mechanism SHM ensures that when there's an F around there tends to be a G around* if

1. FG is hereditary, and consists of a G-mechanism, and an F-mechanism disjunctively recruited for the input of the G-mechanism.
2. There is a specialised homeostatic mechanism SHM that explains that Fs are sufficiently*[7] good indicator of Gs accross generations, and
3. *[The causal mechanism SHM ensuring that when there's an F around there tends to be a G around]* is the HPC that stems from using the instantiations of the properties of *Being an F* and *Being a G* that FG has relied upon as seed[8] and SHM as specialised homeostatic mechanism.

Here we have a mental state whose content involves the presence of a higher order HPC that connects two (lower order) HPCs such that the same cognitive system has, also, mental states whose content involves them[9]. We have used two names for this very HPC: *Q* and *[The causal mechanism SHM ensuring that when there's an F around there tends to be a G around].* The content attributions warranted by etiosemantics are *de re* through and through, so it is *quite* the same to attribute the content

7 It is sufficiently* good only if

$$P(G) \cdot P(F \wedge \neg E|G)(w_{11} - w_{12}) - P(\neg G) \cdot P(F \wedge \neg E|\neg G)(w_{22} - w_{21}) > 0$$

where *E* is the property that makes the G-mechanism fire in the absence of Fs. See appendix B.

8 See the definition of HPC in 1.4.5.

9 For higher and lower order HPCs cf. 1.4.5.

Q exists to FG and to attribute the content *The causal mechanism SHM ensures that when there's an F around there tends to be a G around.* Obviously, though, the latter content attribution is, we may say, "human-readable" while the former is not.

Using the human-readable description of the content is not uncalled for: the state FG with that content is quite literally constituted by a mechanism whose positives have the content *There is an F around*, and another whose positives have the content *There is a G around*. The only material that the disjunctive recruitment is bringing in is that there is a causal mechanism SHM that ensures that when Fs are there Gs tend to be there. This looks, I think, like the kind of fact that a disjunctive recruitment may carry content about.

3.1.2 *Proto-judgements and Proto-beliefs*

FG has a property that we have not encountered before. In the simplest types of mental state we have investigated in previous chapters, those covered by the content-attribution recipe THERE IS AN F AROUND, what was endowed with content was the comparatively short-lived state consisting of some mental mechanism's being *on*. The tokening of that kind of mental state has some affinities with our full-blown *judgements*: it may be suggested, for example, that the going *on* of an F-mechanism bears interesting resemblances with an act of acceptance, on the part of the subject, of the state of affairs the F-mechanism represents. There is no true act and true acceptance, to be sure: the firing of an F-mechanism is hardly an *act* on the part of the possessor of the state, although it is, like acts and unlike standing beliefs, an event. Likewise, the very simple agents we are dealing with can hardly be credited with the ability to *accept* a proposition; but, like those more sophisticated creatures who do accept, they let their behaviour be guided by the content of the mental state it "accepts", if all goes well. The ordinary notion of judgement seems to require a certain amount of rationality (which in turn requires an ability to draw some simple inferences, etc.) absent in the cognitive systems we are studying here.

But at least, in summary, the firing of an F-mechanism is like a judgement in that it is an event of tokening a contenful mental state such that it is supposed to guide the behaviour of the agent. On the other hand, the state with the content *The causal mechanism SHM ensures that when there's an F around there tends to be a G around* has affinities with our standing beliefs: it is not a short-lived but a permanent state of the agent's cognitive system that is to be credited with content. This state is supposed to guide the agent's behaviour in that, when all goes well, makes the agent act in a way that is appropriate to the state of affairs it represents. Henceforth I will help myself to the labels *proto-judgement* and *proto-belief* in cases such as these. Also, when I use terms such as *belief* (or, in chapter 6, *knowledge*), I wish to be taken as talking about these *proto-* counterparts. Which, I think, are just like the real article except for the fact that they are embedded in a cognitive system that does not meet the minimal requirements of rationality, accountability, etc. I will not engage in further discussion of these minimal requirements here, either.

3.2 SECOND-ORDER FUNCTIONS

In the last section we have dealt with states that record relations among real kinds. These states were produced by natural selection: first F- and G-mechanisms emerged. Then, through more natural selection, agents were endowed with disjunctive recruitment among those mechanisms. The states consisting of two mechanisms connected by disjunctive recruitment are the ones that SHM BETWEEN FS AND GS talks about .

But such states as these, hardwired by natural selection, are of limited interest: in fact most of our contentful states are *ephemeral*. That is, they come to exist, and cease existing, during our lifetime and never get to be selected for. In most cases, when we come to associate two real kinds G and F in a way that records a causal link among these properties, it's through a process of learning, and not of brute mutation-cum-selection. While a detailed investigation of the process of conceptual learning is outside the scope of this work, fortunately, teleosemantics -and related approaches, such as mine- provide a way to tackle this problem, so to say, from an abstract implementation-perspective: the main idea (suggested by Millikan (1984)) is that the content of ephemeral states is fixed by facts having to do with selection *up* the implementation ladder[10]. The strategy consists in postulating the existence of a system that has been selected because it gives rise to further states, which are the ones to which content will be attributed. These states have not been selected; their existence is way too short for that. They are, nevertheless, contentful: their existence enjoys a content-conferring biological explanation, through the workings of their producer -which has been selected in the usual way.

Let us return for a moment to the ascription of biological function for devices. A general, simplified template for first-order explanations that ground function-attributions may be[11]:

1ST ORDER FUNCTION: An agent A has a token mechanism, M, with the function FUNCT if the following explains that M exists:

1. The fact that mechanisms of type M have done FUNCT in A's ancestors has been fitness-contributing for them, and
2. This figures in an explanation of the fact that A has M.

1ST ORDER FUNCTION is a schema for the attribution of *first-order* functions because one and the same mechanism is both the bearer of

10 There is at least another possible proposal: that ephemeral beliefs undergo a small-scale process of ontogenetic natural selection that recapitulates the phylogenetic one:

> Suppose our individual psychological developments throw up new possible belief types, new ways of responding mentally to circumstances, at random, analogously to the way that out genetic history throws up mutations at random. Then we would expect such new dispositions to become 'fixed' just in case the belief tokens they give rise to lead to advantageous (that is, psychologically rewarding) actions, analogously to the way that genetic mutations become fixed kust in case they have advantageous (offspring-producting) results. Papineau (1987, p. 66)

There is one case in which this kind of *selection-based learning* -Papineau (2006, p. 186)- can be recovered in the Millikanian terms I favour in the main text: a mechanism MEC creates an environment which fosters this kind of random generation of alternatives, and such process culminates in the emergence of useful mental states; this may explain selection for MEC, and help provide content for the emerged states.
The alternative case -that there is no MEC, and the random generation of alternatives happens also randomly in every generation- is too empirically implausible.

11 Here again, I am ignoring a great many issues that call for complications in the function-attributing recipe. Those issues, though, are irrelevant for my current purposes.

function, and the one that was selected for. When *second order* functions are attributed, these two features come apart: the bearer of function is a *product* of the selected state. A general template for such functions could be:

2ND ORDER FUNCTION: An agent A has a token mechanism, M, with the function FUNCT if the following explains that M exists:

1. There is a mechanism, N, such that, according to 1ST ORDER FUNCTION, N has the function of producing mechanisms of kind K, and
2. N has produced M in A.

It will not have escaped the attention of the reader that M's function FUNCT does not appear in the clauses of the definition. Instead there is an appeal to a kind K of mechanisms. The problem of providing conditions for the attribution of second order functions is, precisely, the problem of characterising K, such that N may have the first order function of creating mechanisms of type K, and such that M counts intuitively as having the function to do FUNCT, in virtue of its being a mechanism of type K.

It is notoriously difficult to do such a thing. For instance, an easy way out will not do. If we postulate that the type K is identical with the type *Mechanisms with the function to do FUNCT*, it is clearly true that mechanisms of such type have the function to do FUNCT in virtue of their being K; but it is also true that we have not explained how N may have the first order function of creating mechanisms of this type.

Another idea is that K should be, simply, *Mechanisms that do FUNCT*. Millikan's proposal of introducing *derived proper functions* could be understood as a sophisticated elaboration of this idea. Unfortunately, even if her proposal is, I think, the best available, I am unsure that it succeeds. I will now review Millikan's theory, and formulate a couple of worries. After that, I will sketch what I take to be a more plausible way out of this difficulty, for the purposes of a theory of content.

3.2.1 *Millikan on Relational, Adapted and Derived Functions*

As I have said, a certain mechanism has a 2^{nd} order function because it has been produced by a mechanism which has the 1^{st} order function of producing mechanisms of its type -see fig. 7. This kind of 1^{st} order functions are, in Millikan's terminology (*cf.* Millikan (1984, chapter 2)), *relational proper functions*. A couple of examples (Millikan's examples) will make clear what is meant by that:

Some chameleons have the ability to make their skin change colour, and this has been evolutionary useful, because it has helped the chameleon hide from danger[12]. Let us assume that a certain mechanism CH in the brain of the chameleon is in charge of issuing the order to the skin chromatofore-cells to rearrange themselves, so as to match the colour of the chameleon's surroundings. The evolutionary history of CH allows 1ST ORDER FUNCTION to warrant an attribution of the

12 Or so goes the folk account of chameleon colour-change. In fact, it has been recently suggested that the evolutionary advantage of colour change has had more to do with "increasingly conspicuous signals used in contests and courtship" -*cf.* Stuart-Fox and Moussalli (2008)- than with camouflage. The folk account is more useful for our present purposes; I hope the lack of biological accuracy will be excused.

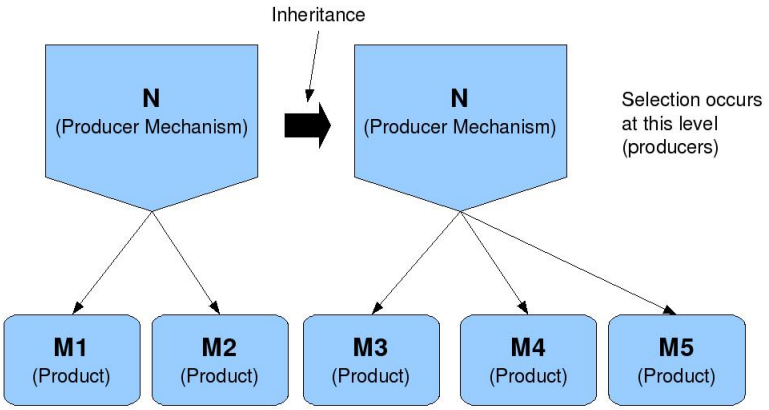

Figure 7: Producer and Products

1st order function of *rearranging chromatofores so as to match the colour of the chameleon's surroundings*.

Bees communicate the position of a source of nectar to the rest of bees in a hive by performing a waggle dance. This has been historically benefitial, as it has lead to better exploitation of those sources. Let us assume that a certain mechanism WD in the brain of the bee is in charge of issuing the motor orders that cause the movements that constitute the dance. The evolutionary story of WD allows 1ST ORDER FUNCTION to warrant an attribution of the 1st order function of *producing dances the pattern of which corresponds in a certain way with the location of nectar.*

The 1st order functions of CH and WD are different to, say, the kidney's function of cleaning blood in that the former do, and the latter does not, have a free slot that must be contributed by the enviroment (for chameleons: by the chameleon's surrounding; for bees: by the location of nectar). The 1st order functions of CH and WD are functions to create a *relatum* (for chameleons: the chromatofore arrangement; for bees: the waggle dance) so that it stands in a predetermined relation (for chameleons: sameness of colour; for bees: one being transformable into the other by the set of rules we may call *B-mese*) to the *relatum* provided by the environment. Thus the name of *relational* proper functions.

More terminology: Millikan proposes to give a name to the relational proper function of a device once in the presence of a particular *relatum*, that is, once saturated the free slot. This will be its *adapted proper function* Millikan (1984, p. 40). For example, if the chameleon's surroundings happen to be brown, CH acquires the adapted proper function of *re-arranging chromatofores so as to look brown*. The product of CH in these circumstances (*i.e.*, the concrete brown arrangement of chromatofores) Millikan calls an *adapted device* (*ibid*.).

Relational and adapted proper functions are 1st order functions: functions attributed to the selected-for mechanism. This is not enough to solve the problem of ephemeral contentful states. For example, we want to provide content attributions for individual bee dances, but

individual bee dances do not have 1st order functions of any kind -they have not been selected for. Now, we have said that contents, according to Millikan's biosemantics, are fixed by the evolutionarily normal conditions for the performance of the function of the consumers of the contentful state. If the contentful state is ephemeral, there are no evolutionarily normal conditions of the right kind. So we need some *other* kind of normal conditions for the ephemeral contentful state to collaborate in the biological function of its consumer. A way to do so may be to provide a proper function for the ephemeral contentful state itself.

Before going on to present Millikan's derived proper functions, I think a caveat is in order. The point made in this paragraph, namely, that providing a function for ephemeral states is necessary in order to provide a content for them, is a common assumption among commenters of Millikan. Thus, *e. g.*,

> Relational and derived functions are essential to account for the capacity to represent something never before encountered in the history of the individual or of the species. Shea (2004, p. 49)

or

> The introduction of derived proper functions goes some way towards solving the problem of content for novel beliefs. Kingsbury (2006, p. 38)

"Representation" is, for Millikan, a technical term that refers only to contentful states of sophisticated cognisers: beliefs and desires are, and waggle dances and chromatofore patterns are not, representations. For the less sophisticated counterpart Millikan reserves the name of "intentional icon". I think the theory of derived proper functions is only designed to deal with representations, and not with contentful states that are just intentional icons. There is some textual evidence for this claim:

> The theory of representations (*as distinct from that of intentional icons*) rests very heavily on the theory of "derived proper functions". Millikan (1997, p. 94; my emphasis)

The most natural reading of this quote is as saying that the theory of representation *only insofar* as it goes beyond the theory of intentional icons does rely on derived proper functions. Her Millikan (2002) provides more evidence, to be discussed below.

So, Shea and Kingsbury, in the quotes above, are strictly right. That is, right if taken to be raising points only about representations such as beliefs, but not about intentional icons. It is likely, though, that they take their points to apply also to intentional icons in general. I will now provide some reasons to think that derived proper functions are, at least, suspect. If I am right, then, a theory which provides contents for ephemeral states without resorting to them will be preferable. As I say, Millikan may be providing just such a theory about simple contentul states (her intentional icons). If so, the problems to be raised against derived proper functions are only a problem to the theory of simple contents that some people take Millikan to be defending; not to Millikan's[13].

13 Of course, my objections *are*, if successful, a problem for Millikan's theory of derived proper functions themselves. I discuss some of Millikan's views on *representations* in 4.5.

So, can we provide a function for the adapted device itself? That is, for the brown arrangement of chromatofores, or the particular waggle dance. According to Millikan,

> The proper functions of adapted devices are derived from proper functions of the devices that produce them that lie *beyond* the production of these adapted devices themselves. I will call the proper functions of adapted devices *derived proper functions*. Millikan (1984, p. 41)

The idea is that a producer mechanism, such as CH or WD, has many functions apart from that of producing individual dances or pigment arrangements. Some examples: making the chameleon invisible, feeding the denizens of the beehive, etc. All of those extra functions of the producer mechanism, Millikan contends, should be considered *derived functions* of its product. Now, among these extra proper functions, producer mechanisms -once we fix the relevant features of the environment, as we have seen- can be credited with *adapted* proper functions that lie beyond the production of the adapted devices themselves: e.g., the adapted proper function of getting bees to go *in thus-and-so a particular direction* looking for nectar; or the adapted proper function of making the chameleon indistinguishable from its *brown* surroundings. These should also be considered as (adapted) *derived proper functions* of the products. I see several problems with the notions of adapted and derived proper functions. Let me present them in turn.

First of all, there appears to be some abuse of language in the talk of adapted proper functions. It is not clear what has made these effects of a device earn the label of (even if adapted) *proper functions*. That is, it is unclear whence comes the normativity in this case. We have accepted that the evolutionary history of a device may endow it with proper functions; but selection, in the case of CH, can only account for its relational functions. Precisely, the idea is that the production of a certain concrete chromatofore arrangement may be completely new in the history of the species, and not selected-for. In the absence of additional arguments, thus, one may be reluctant to honour this set of effects with the name of adapted proper functions. Millikan is aware of this issue:

> [An adapted proper function] is not of course a simple, but only a conditional proper function of the mechanism, an "iffy" proper function, but the "if" part has been asserted. Millikan (2002, section 7)

I read Millikan as endorsing the following argument.

1. N has the relational proper function [for all x, if x do $f(x)$] [*e .g.*, to arrange brown pigment in case the ground is brown].
2. (from 1, and by definition) If a, then N has the adapted proper function to do $f(a)$
3. a
4. (from 2 and 3) N has the adapted proper function to do $f(a)$.

Let me introduce the notation $F_N p$ to mean N *has the proper function to do p*. If Millikan had talked, without qualification, of *proper functions* both in 1 and in 2 she could be accused of ignoring an important scope distinction: that between

- $F_N\,[\forall x\,(x \rightarrow f\,(x))]$ -the relational proper function in 1- which, after elimination of the quantifier, gives $F_N\,[a \rightarrow f\,(a)]$

and

- $a \rightarrow F_N f\,(a)$.

This transition is not in general correct: [producing a key with shape S if presented with a key with shape S] can be the proper function of a key-copier, even if [producing a key with shape S] is not one of its proper functions because, say, the key-copier was not designed to produce keys with that, or any other, shape in particular.

Of course, Millikan is more careful than that, and she qualifies with *relational* and *adapted* the functions in 1 and 2. That is, the transition is from

- $F_N^{rel.}\,[a \rightarrow f\,(a)]$

to

- $a \rightarrow F_N^{adap.} f\,(a)$

Thus, the protest over scope is misplaced: all we have in the transition from 1 to 2 is introduction of terminology. That is, in the argument, stating 2 is *nothing over and above* stating 1.

If so, it is somewhat misleading to detach the consequent of the conditional, in 4 ("an "iffy proper function, but the "if" part has been asserted".) It makes it look as if N had the -admittedly adapted, but also proper- function to do $f\,(a)$. But N does not have the proper function to do $f\,(a)$, not even if *a*. Asserting 4 cannot be anything over and above asserting that

ADAPTED PROPER FUNCTION: $F_N^{adap.} f\,(a) \equiv_{df} \left[a \wedge F_N^{rel.}\,[a \rightarrow f\,(a)]\right]$.

That is: to attribute an adapted proper function is just a way of saying in one breath that the mechanism has a relational proper function, and than one of the *relata* is thus-and-so, or, alternatively, that such-and-such conditions hold. Adapted-proper-function attributions are hybrid statements: on the one hand, they ascribe a relational proper function to a device; on the other, they assert that the environment has such-and-such a property. Talk about adapted proper *functions* is slightly misleading in that adapted-proper-function statements are not solely about functions. Anyway, once we are clear about this, we can use adapted-proper-function talk as a convenient way to abbreviate talk of relational functions when one of the *relata* is fixed[14].

On the other hand, there is the introduction of derived-function talk:

DERIVED PROPER FUNCTION: If N has produced n, then

$$F_N p \equiv_{df} \left[F_n^{der.} p\right].$$

And, by application of DERIVED PROPER FUNCTION in the case of adapted proper functions,

14 And we can take Millikan's word that this is how she intends adapted proper functions to be taken:

> Every reference to an adapted (...) proper function is really an implicit reference to one or more deeper relational functions Millikan (2002, section 7)

$$F_N^{adap.} f(a) \rightarrow F_n^{adap.\,der.} f(a)$$

This result, if understood as dealing simply with transformation of terminology, is pretty harmless and may well be useful in some contexts: saying that a certain ephemeral device has the function of doing $f(a)$ would simply amount to saying that *its producer* has the relational function of [if a, doing $f(a)$], and a. But, while the introduction of adapted-function talk is a simple, innocent terminological move, the introduction of derived function is apparently not. Millikan appears to assume that there is a genuine *transfer of normativity* from producer to product. Thus the contention, in the quote above, that "[t]he proper functions of adapted devices are derived from proper functions of the devices that produce them...". This assertion, unhedged by the likes of "we may say that..." or "it's theoretically fruitful to talk as if...", seems to show that Millikan takes products to have proper functions of their own, with their own charge of normativity[15].

In summary, if the functions of adapted devices are not intended to come for free, as a result of terminological manipulation, but are independent sources of normativity, this I think must be contested. Counterexamples appear to be provided by flawed adapted devices. Consider the following story:

WRONG DANCE! WD is a bee-dance-producing mechanism. It has, as one of its relational proper functions, the production of dances that correspond, under a certain set of transformation rules which we may call B-mese, to locations of nectar around the beehive. Now, let us fix the relevant *relatum* in this relational proper function: a bee has found nectar in the same direction of the Sun, at a somewhat longish distance. It, thus, proceeds to signal these facts to its peers and, to that avail, the WD token in it produces dance M. But, alas, something has gone wrong: M, in B-mese, means "nectar in the *opposite* direction of the Sun, at a somewhat longish distance".

If Millikan is right, the WG in WRONG DANCE! has the adapted proper function of signalling nectar in the same direction of the Sun. This function goes beyond the production of dances, so the adapted device M also has that as an adapted derived function. In this example, M *misfunctions* in indicating the location of the source of nectar. Or does it? Does the performance of M really have this normative dimension?

There is an innocent sense in which it does: the one that makes talk of the function of M a mere notational variant of the talk of the function of WD (see above). But, if we aim for a less innocent sense, in which talk of the function of a product is not simply talk of the function of its producer, we need to produce reasons why this talk is appropriate. Remember -from 1.2.3- that, in the context of a naturalistic program such as Millikan's or mine, we are allowed to indulge in talk of misfunction of a device only because we are able to cash it out as talk of *not doing whatever it is that explains the existence of the device*. And it is simply not true that M exists *because* it has successfully signalled

15 Admittedly, in more recent work (such as the (2002) piece quoted above) Millikan states that derived proper functions are nothing over and above relational proper functions. Even so, the question remains: *whose* relational function? If Millikan means the product's (as opposed to the producer's) relation function, the points I raise below go against this later version of Millikan as well.

the location of a source of nectar. M has just started to exist, and it has never signalled anything before -maybe never before has nectar been in the direction of the Sun at a somewhat longish distance. Can we, alternatively, trace M's malfunctioning to WD's malfunctioning? Well, clearly, WD has not fulfilled its adapted function to signal nectar in the same direction of the Sun. If it had fulfilled it, it would have produced, not M, but a *different* dance M* that does represent that location in B-mese. The only way to turn this fact into a malfunction for M is to say that M has malfunctioned by *failing to be* M*. And it seems quite obviously wrong to put things this way. Things do not malfunction *in virtue of the fact* that they are not other things. The general problem is that etiological accounts of function don't seem to allow for ephemeral devices, that are not part of reproductively established families and are produced in the absence of intentional agents[16], to have function.

Should we conclude that M does not have the possibility of malfunctioning and, therefore, in familiar Wittgenstenian fashion, that it has no function? Millikan has recently suggested a way out:

> [It may seem that a particular bee dance has no proper function], for it seems theoretically possible, at least, that the particular bee dance has no ancestors. (...) Then this particular bee dance, having never occurred in the past, certainly could not have been selected for any effects that it had, hence could not possibly have any proper functions at all. But (...) [w]e must describe functions and how they are performed in the most general way possible. Because bee dances that map different directions are different from one another in specific respects does not mean they are not also the same in more general respects. (...) And when they function in the way that has accounted for the natural selection of their producers (...) they always do exactly the same general thing. They produce a direction of flight that is a given function (mathematical sense) of certain aspects of their form. (...) In this respect, all bee dances of the same bee species have exactly the same proper function. Millikan (2002, p. 130)

The idea is the following: the particular dance, M, *qua* token of the type *dance with such-and-such a meaning in B-mese* may have no function, but M belongs to other types too; in particular it is a token of the type *bee dance*. And bee dances *qua* bee dances have the function of signalling the location of a source of nectar. So, M has no function *qua* dance with such-and-such a meaning, but it has function (and also adapted function, once fixed the relevant *relatum*) *qua* bee dance *simpliciter*.

As far as I can see, this represents a change of doctrine from Millikan (1984): now, we don't need to see the function of a bee dance as *derived* in any sense from the function of its producer. Rather, bee dances themselves, *qua* bee dances, are selected for. They form a reproductively established family in Millikan's sense with a certain character that

16 It seems that intentional agents may produce, say, a prototype model of corkscrew that fails to open bottles. I think an appeal to the intentions of the agent in question -which in turn is susceptible of naturalisation- should be able to account for this fact. It remains to be seen whether this would count as an etiological account of the function of this corkscrew, though.
This is not to be taken as a point about the function of artifacts in general. The function of most artifacts -say, normal corkscrews or chairs- is perfectly naturalisable through the Millikanian theory of proper functions.

correlates with a certain effect etc. Everything behaves according to the definition of proper functions in Millikan (1984, pp. 27-8). They also have a 1^{st} order function according to 1ST ORDER FUNCTION. The apparatus of derived functions, and of 2^{nd} order functions, is doing no real work.

The second problem is more important: as I said at the beginning of the section, an important theoretical advantage teleosemanticists hope to achieve from the postulation of derived functions is the possibility to explain that mental states, concepts and the like, that come to exist in the life of an individual, have content. If content is fixed by the Normal conditions for the consumer of the contentful state to perform its function, as biosemantics has it, ephemeral consumer systems must have function. For the same reason, if two different concepts, both products of the same concept-producing mechanism, are to have different meanings, there must be different Normal conditions for the successful performance of their consumers.

The strategy for adscribing proper functions to ephemeral states that we have just reviewed, on the other hand, would have to be translated to the concept case thus: individual beliefs (or, rather, the consumer systems of those beliefs), *qua* individual beliefs, do not have function; but *qua* beliefs they do. Of course, *qua* beliefs all beliefs will have the same (relational) proper function. Whence comes the difference in meaning then? It has to come from the *adaptor* in each case; the relatum that the world fixes. There must be a mapping function from beliefs to states of affairs that take transformations of ones into transformations of the other. The existence of this mapping function is the Normal condition for the emergence of the functional category of beliefs *qua* beliefs. A full discussion of this contention will have to wait until section 4.5; but it should already be evident that asking for the existence of such a transformation function between beliefs and states of affairs *as a prerequisite* for the contentfulness of said beliefs is asking for quite a lot[17].

3.3 THE ETIOSEMANTIC TAKE ON EPHEMERAL CONTENTFUL STATES

The conclusion of the last section is that the (derived, 2^{nd} order) function of ephemeral contentful states does not play any substantial role in fixing their content. In fact, a stronger claim might be made: it is unclear that there are derived, 2^{nd} order functions at all. If so, a powerful reason to recommend etiosemantics against classical teleosemantics is that it can attribute contents to ephemeral states without the need of positing derived functions. This section shows how it goes about doing that.

In the first two chapters (see, particularly, 1.3.2 and 2.2.2) I have given my reasons to think that relying on 1^{st} order functions to fix content, in the way teleosemantics in general and biosemantics in particular do, makes content attributions vulnerable to the Input, Output or Indeterminacy Problems. I have also defended that letting content be fixed by the natural structures I have called HPCs patches this vulnerability. The way to make the same idea work for what I've been calling ephemeral states is by relying on higher order HPCs. Suppose that we wish to provide content attributions such as the ones

17 For additional criticism of the notion of derived function, see Preston (1998).

warranted by SHM BETWEEN FS AND GS, but for states that have not been selected for; only produced by a certain selected-for mechanism N. What we need is an HPC that connects cues of a certain type C -the ones that N will use to effect disjunctive recruitments- with lower level HPCs of the kind SHM BETWEEN FS AND GS relied upon; of course, this time the homeostatic mechanism that individuate these HPCs may be much more short-lived that the ones relied upon in SHM BETWEEN FS AND GS -those needed to stay around for as long as selection for the mental state took; these need simply make a disjunctive recruitment fitness-conducive during the lifetime of an individual mental state.

We want to tell the following story about the appearance of an ephemeral mental state with a content analogous to *The causal mechanism SHM ensures that when there's an F around there tends to be a G around*:

> M exists in A because, many generations ago, a mechanism N appeared in one of A's ancestors -through random mutation, we may suppose. N has properties of type C as input and, when it goes on, it effects a disjunctive recruiting among two pre-existing brain mechanisms that produce contentful states; which ones exactly is a function (in the mathematical sense) of C; say, if $f_1(C) = F$ and $f_2(C) = G$, the recruitment is produced between F- and G-mechanism.
>
> In many occasions, the recruiting effected by N has increased the overall fitness contribution of the two mental states. As a consequence of this, N has become fixated in A's population.
>
> Besides, there is a higher order HPC that connects properties of type C with the HPCs that have made disjunctive recruitment between $f_1(C)$-mechanism and $f_2(C)$-mechanism fitness contributing in the process leading to fixation of N.
>
> Today, a cue C′ has caused A's token of N to produce a disjunctive recruiting of A's F-mechanism by A's G-mechanism, giving rise to M (cf. figure 8.)

The content of M in this case is fixed by the lower level HPC that goes together with C′ in virtue of the homeostatic mechanism of the higher level HPC. Although it may sound convoluted, the content-attributing recipe is, in fact, straightforward. Let me give another pass to this HPC hierarchy:

- There is a higher level HPC (let us call it Q^n; the superindex stands for the order of the HPC) that connects (for as long as needed for N to become fixated) each of a number of cues C_i with a lower level HPC Q_i^{n-1}.
- In their turn, repeatedly in the process of selection for N, HPCs such as Q_i^{n-1} have kept together the properties needed to make a disjunctive recruitment between $f_1(C_i)$-mechanism and $f_2(C_i)$-mechanism fitness conducive.
- If a cue C_k causes N to produce a new disjunctive recruitment among a $f_1(C_k)$-mechanism and a $f_2(C_k)$-mechanism, we say that the content of the new (ephemeral) disjunctive recruitment is Q_k^{n-1} *exists* -or, in the human-readable version, *The causal mechanism* SHM_k^{n-1} *ensures that when there's an* $f_2(C_k)$ *around there tends to be a* $f_1(C_k)$ *around.*

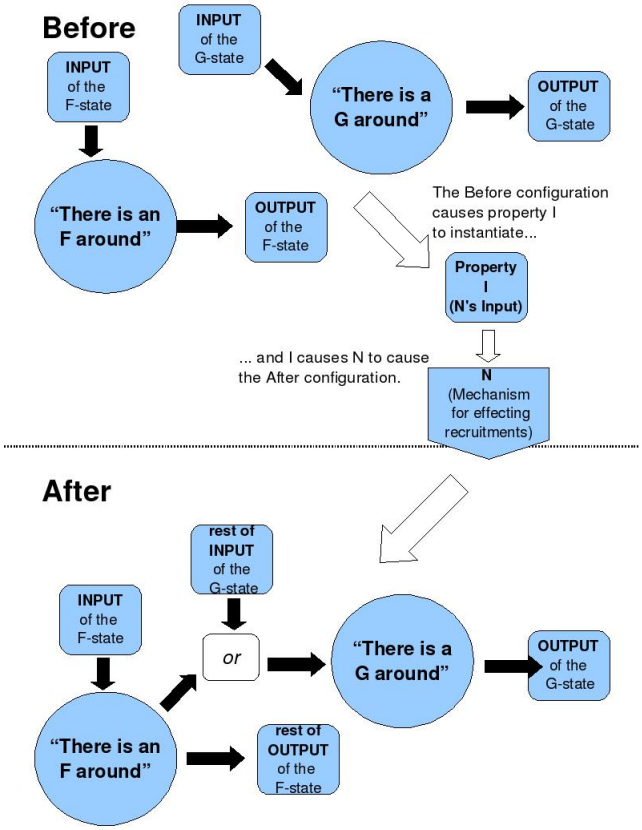

Figure 8: The Explanation of Ontogenic Appearance of Recruitments

We may even provide a general process to devise recipes for attributing content involving an HPC to ephemeral states[18].

EPHEMERAL CONTENTFUL STATE: To devise a recipe for attributing ephemeral states with the content $F(k)$ where *k* is an HPC:

A1: Find a recipe to attribute contents of the type $F(k)$ to a *selected-for* state S.

A2: Find out in which way does S's causal profile's being fitness-conducive depends on the Property Cluster of kind *k*.

Once done so, we may provide the following recipe. Consider a mechanism N that is caused, by cues of type C, to create states whose causal profile's being fitness-conducive depends (on the way found out in A2) on the Property Cluster of an HPC Q_i^{n-1} that is function of the cue: $Q_i^{n-1} = f(C_i)$. Now, if

B1: N's having created states from cues C_i have made it fitness conducive in many of the possessors of N, contributing to its fixation in their population of its possessor,

B2: There is a higher order HPC Q^n that explains that instantiations of properties C_i go together with the instantiations of the properties in the HPC $Q_i^{n-1} = f(C_i)$, and these lower level HPCs in turn explain the fitness conduciveness alluded to in B1.

Then the state created by N because of cue C_k has as content $F\left(Q_k^{n-1}\right)$.

It will be very much easier to see how this process works after discussing a concrete example. In the following subsection I consider long term potentiation, which could well be a real life producer of ephemeral contentful states.

3.4 LONG TERM POTENTIATION

Long term potentiation (LTP), is a brain process which

> results from coincident activity of pre- and post-synaptic elements, bringing about a facilitation of chemical transmission that lasts for hours *in vitro*, and that can persist for periods of weeks or months *in vivo*. Cooke and Bliss (2006, p. 1659)

The mechanism of LTP, widely perceived as a possible implementation basis for memory and learning (*cf.* Koch (1999, p. 317)) is probably a process "up the implementation ladder" of the kind we are interested in. It is, under the plausible assumption that LTP exists because frequent coincident activity of neurons is not, well, *coincidental*. If such activity responds, instead, to a causal relation among the environmental properties to which the pre- and post-synaptic neurons are sensible, it makes sense to increase the degree to which the pre-synaptic neuron

18 It cannot be a watertight, automatic process, though: the circumstances for different kinds of contents will be different, and some thinking on the part of the theorist will be needed. This is fine from a naturalistic point of view, though: the thinking in question will have to be done only a finite number of times; only, that is, for the mechanisms that create the ephemeral contentful mental states that are human thoughts.

makes the post-synaptic neuron fire -in the simple model we have been working with, it makes sense to disjunctively-recruit pre-synaptic firing for the input of the post-synaptic neuron.

To keep the discussion in focus, let us suppose that an agent A has a mechanism we could call LTP* (intended to be a simplified, manageable version of the LTP mechanism) that could be described as follows:

LTP*: Whenever two mental mechanisms, an F-mechanism and a G-mechanism, are *on* together repeatedly, LTP* builds a probabilistic causal connection that makes the G-mechanism fire with a higher probability if the F-mechanism has fired, and vice versa. This probabilistic causal connection decays slowly when there is no correlated activity between F- and G-mechanism.

LTP* is our candidate for a mechanism N that produces ephemeral contentful states. Here, the cues C_k are of the type *An* F_i*- and an* F_j*-mechanism being on together repeatedly*. Connecting probabilistically both mechanisms in the presence of this cue has been fitness contributing in the past because, many times, whenever both mechanisms are on together repeatedly, there was a causal relation between *there being an* F_i *around* and *there being an* F_j *around* that made both states of affairs be positively correlated. Building up the causal connection among states ensured that the system as a whole would miss less instantiations of the states of affairs in question.

Now, for LTP* to be able to confer content on its productions, it must be the case that there is a higher order HPC connecting instantiations of the cues C_k with the causal grounds connecting the states of affairs *there being an* F_i *around* and *there being an* F_j *around.* That is, there must be a causal explanation of the following state of affairs:

ASSOCIATION: Many times[19], when there is frequent cofiring of an F-mechanism and a G-mechanism in S's brain, there is a causally grounded relation between there being an F around S and there being a G around S.

And so, previous coinstantiation making it probable that there is a grounded relation between F and G, *future* coinstantiation is to be expected[20]. See Appendix C for further discussion.

While it is clear than in many spatiotemporal domains around us ASSOCIATION is causally grounded, it is not trivial to explain exactly what are these causal grounds. There are so many properties instantiated at so many different times and places that, in general, correlation among property instantiations is no guarantee of causal relationship. ASSOCIATION is only plausible for creatures that have states attuned to the instantiation of *some* properties, under *some* circumstances. That is, I dare to say, it is unlikely that a mere associative mechanism such as long term potentiation, on its own, would have given rise to contentful states. Anyway, suppose that we have identified the domain under which

19 This "many times", again, is susceptible of rigorous analysis in terms of Fitness Contributions.

20 Could we have proposed the following alternative explanation to HM?

(HM*) Whenever there is frequent coinstantiation of two mental states, F-mechanism and G-mechanism, it is sufficiently probable that there is a grounded relation between F-mechanism and G-mechanism.

This "explanation" makes LTP* redundant. If what explains the coinstantiation of mental states is that they are groundedly related, this does not explain (but actually preempts) the usefulness of a mechanism that builds a grounded relation between them.

ASSOCIATION is causally grounded, and which has helped selection for LTP*. We may then provide a content for the products of LTP*:

LTP* PRODUCTS: If there is a higher order HPC Q^n that explains ASSOCIATION, then the state created by the token of LTP* in S, as a result of the cue C_k: *[An F_i- and an F_j-mechanism being on together repeatedly]* has as content *The causal relation R ensures that whenever an an F_j is around, an F_i tends to be around.* Where [the causal relation R ensuring that whenever an an F_j is around, an F_i tends to be around] is the HPC Q^{n-1}_k that Q^n keeps together with the cue C_k.

What explains that LTP* has linked F- and G-mechanisms together? The answer to this is that, a sufficient number of times, the G-mechanism has been *on* whenever the F-mechanism was *on*, and LTP* builds a state whenever this condition holds. And why is that? Because most of the time that two states behave like that, there is a causal relation linking the properties they refer to in the right way. The state created by LTP* means that such a causal relation is in place.

Once we have a mechanism such as LTP* up and running and creating contentful states, content is rapidly divorced from usefulness. The content created in presence of a certain cue is fixed by whatever HPC is kept together with said cue by the higher-order HPC that has enabled selection for LTP*. But this lower-order HPC may be perfectly useless, recording a causal relation between two properties whose connection will never be critical, or even particularly useful, for anybody. In those cases LTP* will have created a useless proto-belief.

3.5 A STEP TOWARDS PREDICATION

This is a good point to stop and discuss a couple of general features of the etiosemantic account of content. We may now, for instance, outline the general process that we have used in devising the content recipes THERE IS AN F AROUND and SHM BETWEEN FS AND GS. In order to build a recipe to attribute contents that are substitution of a certain schema $F(x)$ -such as *There is an x around*- the first thing one needs is a metaphysical account of what it is to be a case in which $F(x)$. Thus, for *There is an F around S*, the underlying metaphysical picture is one in which the state of affairs consisting of an F being around S is constituted by some of the instantiations in the Property Cluster of F (which is an HPC) happening around S (at the time the contentful state is tokened). And for *The causal mechanism SHM ensures that whenever an F is around, a G tends to be around*, the picture is one in which a higher order HPC keeps together the properties *Being an F* and *Being a G*.

Once we have such a rough and ready metaphysical picture of what it is to be the fact $F(x)$, the task of finding a recipe to attribute the content $F(x)$ to a selected-for state reduces to this other task: finding the causal profile of a state such that a complete explanation of the selection for that state must invoke $F(x)$, in a sufficient number of generations. This is the barest outline of the process; a much more detailed account may be gathered from the recipes we have offered for concrete contents, and much more detail still will be provided in the chapters to come.

So, we may think of the process of providing content recipes as a process in which one uses a metaphysical picture of the target fact to ascertain the causal profile of *which* mental mechanisms may have its

selection enabled by such fact. Thinking of it this way suggests that facts whose metaphysical profiles are close may be such that the mechanisms whose positives have them as content also have causal profiles that are close. This may provide an easy route for the appearance of contentful states involving more sophisticated states of affairs. What I have in mind is that the states which may be attributed with content according with SHM BETWEEN FS AND GS may be a precursor of the more interesting relation of predication.

For example, it is possible that, in some cases, the causal structure *SHM* that the disjunctive recruitment exploits is particularly intimate. It may be, for example, that condition 2. in SHM BETWEEN FS AND GS holds because the specialised homeostatic mechanism that subserves kind F is a proper part of the specialised homeostatic mechanism that subserves kind G, and the G Property Cluster is a subset of the F Property Cluster. In such a case, in some sense, being a G is a part of what it is to be an F: G reappearing is part of what it is for F to reappear. This is the kind of relation that holds between pairs of HPC such as *man* and *animal*. In case this more stringent relation holds between Fs and Gs, we may attribute a content that involves being in the determinate-determinable relation.

GS ARE FS: A subject S has a mental state FG with the content *Gs are Fs* if

1. FG is hereditary, and consists of a G-mechanism, and an F-mechanism disjunctively recruited for the input of the G-mechanism.
2. The specialised homeostatic mechanism SHM_F that partially constitutes the HPC F is a proper part of the specialised homeostatic mechanism SHM_G that partially constitutes G. Likewise the G Property Cluster is a subset of the F Property Cluster
3. The facts about F and G cited in 2. are the ones that have enabled selection for FG.

This is a good starting point towards predication. If we abuse language for the sake of (slightly) gaining psychological plausibility, we may render the content of the states alluded to inGS ARE FS thus: *G is F* (*Fly is moving thing*, *Peach is fruit*, etc.)[21]

3.6 EXPLANATIONS AND PROCEDURES

There is a source of disatisfaction here, though: while SHM BETWEEN FS AND GS appears to be identifying a natural grouping of mental states -those that record a grounded relation between the exemplifications of two kinds-, GS ARE FS appears to be tracing an artificial subdivision inside that group, with the only objective of identyfing a relation among properties (the determinable-determinate relation) that *we* human cognisers do find natural. At first sight, that is, SHM BETWEEN FS AND GS points at a theoretically fruitful subclass of mental states, and GS ARE FS does not. This metatheoretic fact is evidence that tracking true necessitation relations among kinds has a more natural home in more sophisticated cognisers. Simple agents such as the

21 This language abuse is inspired by Millikan (in Millikan (1998)), after Quine's (in Quine (1960)) "Hello! More Mama", in her rendering of the content of something similar to what I have been calling F-mechanisms.

ones we have been dealing with have enough with tracking grounded relations of sufficiently* increased posterior probabilities -although if the grounded relations in question are of the kind recorded in GS ARE FS, they will end up having states with the content *Gs are Fs*.

This talk of theoretical fruitfulness is extremely vague. There is maybe a clearer reason why the recognition of causally-grounded relations appears to belong more naturally with our simple agents: the mechanism of disjunctive recruiting a mental states' firings whenever such recruiting is fitness-conducive does not distinguish between causally-grounded relations and full-blown natural kind parthood relations as recorded by GS ARE FS. So, in a sense, an agent relying on this mechanism cannot be *blamed* for treating equally both kinds of relations. There is a closely analogous sense in which Democritus cannot be blamed for mistaking a black speck for a fly. The procedure he follows for recognizing flies has this sorry feature. I submit, and will try to elaborate on this in section 3.10, that the difference between explanations (that ground content) and procedures (that ground holding-accountable practices) can explain many of our internalist intuitions both in epistemology and in philosophy of language. Although a fuller discussion will have to wait until then, we can already sketch what I will be meaning by *Procedure*.

Procedures depend on the mechanisms that create contentful states. Every such state has a producer that has created it with such and such a causal profile. In EPHEMERAL CONTENTFUL STATE I suggested that, for a mechanism N to be able to confer content on its ephemeral products, it must create them with a certain causal profile (that is, maybe, function of the cue that has caused N to create them). This is what gives meaning to the notion of procedure[22]:

PROCEDURE: If a cue C_m has caused a mechanism N to create an ephemeral state M that can be attributed with content by using a recipe derived from EPHEMERAL CONTENTFUL STATE -in particular, N has historically relied on a higher order HPC Q^n, as described in condition B1-, then

1. M's *Procedure* is the causal profile with which N has created it.
2. M's Procedure's *ecologically-fixed context* is the domain *d* in which the specialised homeostatic mechanism that partly individuates Q^{n-1}_m is active (see subsection 1.4.5).

The hypothesis about blamelessness is, then,

BLAMELESSNESS: M's behaviour is blameless if and only if it accords with M's Procedure. Otherwise, it is blameful.

In 3.10 we will see that this notion of blamelessness[23] allows us to capture many of our internalistic intuitions about the behaviour of contentful states. The idea is that it should do (*should* from this blame-

22 This notion is related to what Millikan (1984, chapter 9) calls *intension*.

23 This expression, introduced as a technical term in Parfit (1984), has become a term of art in ethics, referring to the possibility of acting for the right motives -even if with the wrong outcome. Although I'll put it to a different use, I think there are important, non-accidental analogies between blameless wrongdoing in ethics and in the philosophy of mind. I hope to develop this idea sometime.

ful/blameless perspective; there are others) whatever it was created for doing[24][25].

I turn now to the discussion of another, very important family of ephemeral contents: those involving individuals.

3.7 REFERENCE TO INDIVIDUALS IS COGNITIVELY CHEAP

There is a sense in which the conditions for existence of contentful states which involve reference to individuals are not terribly more complicated than the conditions for states which involve reference to natural kinds. In fact, the very same kind of process through which, I have argued, a mental state is endowed with the content that *There is an F around*, may underlie contents such as *a is around*, where *a* stands for an individual. Plant phototropism is a simple case in point.

Phototropic plants, such as the snow buttercup, turn around to face the Sun. The heliotropic mechanism of buttercups (cf. Sherry and Galen (1998)) involves differential cell-growth in differently-illuminated regions of the flower's peduncle. This, in turn, makes the flower bend towards the source of light. Let us consider a peduncle, PED, which can be in a number of states, PED_i, each of them a substitution of the schema: *Groups of opposed epidermal cells in the peduncle along vector i's having different size*. Each of PED_i indicate a number of properties: *such-and-such pattern of light density in the surface of the peduncle*, *the Sun being in such-and-such position in the sky*, etc.

For each such property we may construct an Indication Profile. In fact, given that we have now many more states than two, this Indication Profile will have to be a hypermatrix. For simplicity, we may consider a finite number of different states PED may be in -say, *n*-, and *n* different apparent positions of the Sun in the sky. Then calculate conditional probabilities for each pair of values -they will maybe be normally distributed around some value, the one that makes the flower face the Sun.

Identifying the Fitness (Hyper-)Matrix for each such IP implies doing empirical research on the beneficial effects of phototropism for the buttercup. According to Sherry and Galen (1998, p. 984), they include: increasing the temperature of the flower bowl, being more attractive for pollinator insects, helping to grow more, and larger, achenes, which in turn are better at germinating, etc[26].

Let us see now whether the situation with snow buttercups warrants an attribution of content to any of PED_i. If we remember the recipe for content-attribution we proposed in Chapter 1:

---

24 It should be noticed that this technical notion of *blamelessness* is totally independent from *responsibility*: these very simple mental states cannot help but behave according to their procedure. Responsibility will only enter the picture once free decisions do, and this is not a matter I plan to discuss in this work.

25 The notion of Procedure is what is needed to account for the possibility of supernormal stimuli, briefly discussed in 1.1.2. Although in our examples we are only seeing Procedures which prompt a mechanism to go *on* or *off*, we could easily imagine more nuanced Procedures which tailored the degree of response to some features of a cue. Such happens with the size of an egg (cue) and degree of preference for incubation (response). All of this is compatible with the content of the representation being *There is an egg here*.

26 Of course, snow buttercups can only cash these benefits if, apart from tracking where the Sun is, they follow it. In very simple systems such as this (what Millikan (1995) has called *pushmi-pullyu* systems), the *descriptive* (*e.g.* telling where the Sun is) and the *directive* (*e.g.* making the flower bend in that direction) functions cannot be separated, so we are helping ourselves to an idealisation here in considering only the descriptive part.

THERE IS AN F AROUND: M's being *on* has the content *There is an F around* if

TFA1: ETIOLOGICAL FUNCTION warrants, for a number $i \geqslant 1$ of properties G_i, the attribution to M of the etiological function of indicating the instantiation of G_i around its possessor S.

TFA2: The different pairs, $\langle IP_{G_i}, FM_{G_i} \rangle$, of Indication Profile and Fitness Matrix for each property G_i are grounded on the frequent co-instantiation of several properties in S's environment. (cf. subsections 1.4.1 and 1.4.2)

TFA3: The fact that $\langle IP_{G_i}, FM_{G_i} \rangle$ pairs remain the same across S's lineage is causally grounded on a specialised homeostatic mechanism *SHM* that explains the recurrence of the properties appealed to in TFA2 in a certain domain *d* around M.

TFA4: F is the natural kind individuated by *SHM* and the smallest set of properties *P′* such that *SHM* explains the fact that, in *d*, the properties in *P′* are frequently coinstantiated.

The properties that clause TFA2 talks about include the ones we have been describing above: differential illumination of the peduncle, increased flower-bowl temperature, etc. We may furthermore assume that the Fitness Contribution of phototropism has been instrumental in the survival of the actual snow buttercups. Now, what is the specialised homeostatic mechanism that explains the recurrence of these properties? It is the mechanism that keeps a light source above the heads of the buttercups: the hydrostatic equilibrium between gravitational compression and fusion that keeps main-sequence stars such as the Sun where they are. The HPC defined by the particular instantiations of the properties above and this homeostatic mechanism is, then, an individual: the Sun. The content of a particular arrangement of cell sizes in a snow buttercup's peduncle has a content along the lines of *The Sun is over there.*

3.7.1 *Natural Kinds And Individuals Are Not All That Different*

The reason that reference to individuals is as cognitively cheap as reference to natural kinds, under this account, is that, fundamentally, individuals are taken to be the same kind of stuff as natural kinds: the same kind of features that identify kinds (i.e., instantiations of properties in a cluster and the causal link among them) identify individuals as well. Individuals, in this sense, are a subset of kinds which meet further restrictions. Thus, *e. g.*, in Wilson et al. (forthcoming):

> However individuals and historical entities are specified precisely, they have in common the idea of being spatiotemporally bounded, continuous particulars.

Individuals are different from *other* kinds in that their spatio-temporal parts are in contact with one another. Kinds in general, on the contrary, may have and normally have scattered parts (*e. g.*, the different tiger instances).

The idea that individuals and kinds are fundamentally the same kind of entities is by no means new, and has been defended by Millikan (Millikan (2000, paragraph 2.3) and Millikan (1998)). A number of

philosophers of biology (starting with the influential Ghiselin (1974)), on the other hand, have defended that biological species are best seen as individuals. Thus the cognitive cheapness: the content-conferring processes we are investigating simply latch on whatever HPC is around. If it is a kind, a kind; if it is an individual, an individual.

3.7.2 *Reference to Individuals is Cognitively Expensive*

On the other hand it is a contingent, but real enough, characteristic of most individuals that they are not longevous enough to sustain the evolution of contentful states with them figuring in the content attributed. It is not by chance that the individual doing this job in the previous section is the Sun. Not that there are no other examples -maybe some trees are longevous enough, and evolutionary biologists are probably able to grow in chemostats bacteria that are sensible in interesting ways to short-lived concrete individuals[27]. But the point remains that individual lifespans are, in general, much shorter than the time needed to sustain the selection for a contentful state.

Contentful states with contents involving individuals will be, in general, ephemeral. In this section I will provide the template for content-attributing recipes for these ephemeral states.

3.8 INDIVIDUALS AND EPHEMERAL STATES

What we need for the existence of ephemeral states with a content involving individuals is a selected-for mechanism (or chain of mechanisms) that produces them. As we have seen in 3.2, the newly born contentful state will (or at least may) be functionless; it is features about the explanation of its existence -and its contribution to the explanation of the existence of its producer- that transcend those needed to fix a function attribution, that fix its content. We may start by providing a set of sufficient conditions for crediting an ephemeral state with the content *a is around*. For that, we can follow the process recommended in EPHEMERAL CONTENTFUL STATE (see 3.3). According to this process, we first need (according to A1[28]) to obtain a recipe for attributing a content of the same kind to a selected-for state. As we have just seen in 3.7, the THERE IS AN F AROUND recipe is sufficient for this. We need then (as per A2) to check in *which* way such selected-for state's causal profile depends on the individual's HPC for it to be fitness-conducive.

Let us say, in general, that the selected-for mechanism N that produces the ephemeral mechanisms whose being *on* has the content *a is around* acts according to the following routine:

ROUTINE: Inputs of type C, such as C_m, cause N to create mechanisms with input $f_{input}(C_m)$ and output $f_{output}(C_m)$.

That is, we are supposing that there is a transformation that takes every cue to a causal profile, and N has stumbled upon a way to make a mechanism with the latter causal profile whenever it receives the right cue as input. And the ephemeral mechanisms with the causal profile in question have been fitness-conducive because (here we need to plug

27 And if Ghiselin's "radical solution" of considering species as individuals is right, there are many more examples.

28 This A1, and the A2, B1 and B2 below, are from EPHEMERAL CONTENTFUL STATE.

in the content attribution recipe THERE IS AN F AROUND, which, we have said, will do for *a is around* as well):

FITNESS CONDUCIVENESS: A sufficient* number of mechanisms M_i produced by N according to ROUTINE in the presence of cue C_i have been fitness-conducive times because

- FC1: For a number $j \geqslant 1$ of properties G_j, M_i has indicated the instantiation of G_j around its possessor.
- FC2: The different pairs, $\left\langle IP_{G_j}, FM_{G_j} \right\rangle$, of Indication Profile and Fitness Matrix for each property G_j are grounded on the frequent co-instantiation of several properties in the environment of M_i's possessor.
- FC3: The fact that $\left\langle IP_{G_i}, FM_{G_i} \right\rangle$ pairs remain the same during M_i's lifetime is causally grounded on a specialised homeostatic mechanism SHM_i that explains the recurrence of the properties appealed to in FC2 in a certain domain d_i around M_i.
- FC4: a_i is the individual individuated by SHM_i and the smallest set of properties P'_i such that SHM_i explains the fact that, in d_i, the properties in P'_i are frequently coinstantiated.

Finally (as per B1 and B2) we need to make sure that there is a higher order HPC that keeps together the individuals a_i appealed to in FITNESS CONDUCIVENESS and the cues C_i. The only extra bit we need for this is part of B2:

HIGHER ORDER HPC: There is a higher order HPC Q^n that explains that instantiations of properties C_i go together with instantiations of the properties in the Cluster of HPC a_i.

That is, N (the selected-for mechanism) relies on HPC Q^n in the sense that the cues it uses to create ephemeral mechanisms M_i go usually together with lower level HPCs (individuals, in this example) that are situated so as to make M_i, with the causal profile it has, fitness-conducive. This going together[29] is enforced by Q^n. If these conditions are in place, the content-attributing recipe for ephemeral states is easy:

A IS AROUND - EPHEMERAL: An agent A has a mechanism, M_i, whose positives have the content a_i *is around*, if cue C_i has caused N to create it and all of ROUTINE, FITNESS CONDUCIVENESS and HIGHER ORDER HPC are in place.

This is just an application of the process suggested in EPHEMERAL CONTENTFUL STATE to the case of ephemeral states with contents of the kind *a is around.* The rigorous formulation is, to be sure, somewhat convoluted, but the underlying idea is extremely simple: the selected-for mechanism relies on an HPC (for this purpose: an HPC is a correlation and its causal explanation) that correlates cues and lower level HPCs. If so, the produced, ephemeral mechanism may be attributed with a content that involves the lower level HPC (an individual, maybe, or a kind) that goes together with the cue that caused its creation.

Again, it will be much easier to secure a firmer grasp of A IS AROUND - EPHEMERAL and its auxiliary principles by applying them to a concrete example.

29 *Going together* is metaphorical, but the main idea should be reasonably clear. A cue goes together with an HPC if cues of that type are reliably accompanied by the existence of HPCs of that type.

3.9 DOMINANCE HIERARCHIES AMONG LOBSTERS

American lobsters have a complicated social structure, with males forming a dominance hierarchy -*cf*. Karavanich and Atema (1998). If you put a bunch of lobsters which have never met before in a water tank, males will start fighting with one another. After fighting, losers will subsequently avoid contact with winners, while winners will continue to display aggresive behaviour against losers. According to Karavanich and Atema (1998), the cue lobsters use to avoid fights with previously encountered conspecifics is the chemical signature of their urine.

A natural way to describe the scenario in intentional terms is that an individual lobster (say, lobster #i, or L_i) after a fight with some other individual (say, L_n), acquires a mechanism the firing of which has the content *L_n is around*, which it may subsequently use to avoid further encounters with the winning lobster, or to more efficient harassment of losers. Let us see how etiosemantics goes about predicting this outcome.

We postulate the existence of a mechanism N which follows this routine:

> Whenever L_i loses in a fight with some other lobster, say L_j, and the urine chemical signature UCS_j is around L_i when this happens, N creates a state M_{ij} in L_i with the following causal profile: the presence of urine chemical signature UCS_j around L_i causes M_{ij} to go *on*, and M_{ij} 's going *on* causes L_i to initiate a fight-avoidance routine.
>
> On the other hand, whenever L_i wins in a fight with some other lobster, say L_k, and the urine chemical signature UCS_k is around L_i when this happens, N creates a state M_{ik} in L_i with the following causal profile: the presence of urine chemical signature UCS_k around L_i causes M_{ik} to go *on*, and M_{ik} 's going *on* causes L_i to display aggresive behaviour towards L_k.

This is, simply, the ROUTINE schema we saw above with the following substitutions of variables:

- C_m is *presence of chemical UCS_m together with an outcome in fight O* (which can be either *win* or *lose*)
- $f_{input}(C_m)$ is *presence of the chemical that is part of C_m*.
- $f_{ouput}(C_m)$ is *Initiate avoiding behaviour* if O is *lose* and *Display aggresive behaviour* if O is *win.*

Given how we have set up the case, it is natural to suppose that N is still around because it has been fitness-conducive[30]; and such fitness-conduciveness stems from the fitness-conduciveness of a sufficient number of its products which is, in turn, explained by the fact that:

> Mechanisms such as M_{im} have indicated a number of properties -it indicates the presence of chemical UCS_m very well, and many other properties not that well. Indicating these properties is coupled with certain Fitness Matrices; this fact is explained by the frequent coinstantiation of a number of properties; most relevantly, a certain chemical signature of

30 Apparently, it is still still unclear what it is about dominance hierarchies among lobsters that makes them selected-for. I will take for granted that they *are* selected-for.

> the urine correlates with whatever it is that makes a lobster likely to win or lose in a confrontation with L_i. There is a specialised homeostatic mechanism SHM_m that makes this properties coocur. SHM_m, together with the Cluster that stems from the properties we have just talked about, constitute an individual lobster, L_m.

This is FITNESS CONDUCIVENESS, applied to the lobster example. The Property Cluster of an individual lobster L_m includes: instantiations of whatever signs of lobsterhood you may think of, and of particular properties of that very lobster, such as being particularly fond of some type of frozen squid, having clasps of thus-and-so shape and, crucially, releasing UCS_m with the urine and being likely to win (lose) in a fight with L_i. SHM_m is the homeostatic mechanism that keeps those properties recurring together: the mechanism that prevents $L_m$from disintegrating -the cohesion of the materials of her carcass, the nutritional processes that keep her body from diminishing in size, etc. Finally, it needs to be the case that a non-accidental regularity keeps the C_m being good enough cues of individual lobsters:

> There is a higher level HPC that explains that instantiations of properties of kind C go together with [the fact that a certain urine chemical signature goes together with a likely outcome of fight, for a reason that has to do with an individual lobster].

And this is HIGHER ORDER HPC as applied to the lobster example. The higher level HPC required indeed exists. It is, basically, lobsterhood plus facts having to do with the number of possible urine chemical signatures: if a lobster finds a chemical signature; this is a sufficiently good sign that this encounter upon UCS_m will be caused by the same thing that caused previous encounters, because the probability of two lobsters sharing UCS_m is vanishingly low, and the probability of UCS_m not coming from a lobster is equally low[31]. Supposing all of this causal structure in place we may, finally, provide a content attribution recipe for individual mechanisms M_{ij}:

> A lobster L_i has a mechanism, M_{ij}, whose positives have the content *L_i is around*, if cue C_i has caused N to create M_{ij} (*cf.* figure 9).

Even with the benefit of concrete examples, the appeal to a nested structure of HPCs, closely knitted together with the causal profile of the selected-for mechanism and the ephemeral mechanisms it creates, may strike many as unnecessarily complicated. It is, in fact, strictly necessary if we are to avoid Indeterminacy, Input and Output problems in our content attributions. First of all, the low order HPCs (each individual lobster) are needed to avoid having content indeterminacy among the many properties each ephemeral mechanism indicates (*e. g.*, such and such UCS, this individual lobster, lobsters in general, etc.), or among the many properties that explain the success of the consumers of the representation (*e. g.*, lobster likely to beat (lose against) me in a fight, creature likely to harm (be harmed by) me, etc.). This is simply the ephemeral implementation of the strategy described in chapter 1.

31 Of course, only in the ecologically-fixed context of N's Procedure. See above 3.6 and below 3.10.

Secondly, we need the higher order HPC N relies on to fix the lower level HPC each new ephemeral mechanism is about[32].

I should also point out that I am not conjuring this complicated structure of overarching HPCs out of the blue to serve my content-attributing needs. Those HPCs are already sitting there, for the taking. The content theorist only needs to help himself to them.

Although I have used the lobster case merely as an example in our way to building more complex contentful states, there are lessons to draw from it that may be useful for the contemporary debate about Individual Recognition in ethology. I discuss these implications in Appendix C.

3.10 CONTENTFUL STATES AND NON-EXISTENCE

It should be noted that this general strategy for attributing content to ephemeral states complies with what I called in 1.6 COMPRESSED EXPLANATION: the content of these states is a compressed explanation of their existence in a sufficient number of cases. The difference with contents attributed using recipes such as THERE IS AN F AROUND is that, in the latter, selected-for states, the cases in which the content of the representation is *not* a compressed explanation of its existence are cases in which the representation is, simply, false. But ephemeral states may go wrong in a more catastrophic manner.

Consider again the lobster case. There is a mild kind of misbehaviour on the part of a contentful mental state: If L_i loses a fight with a UCS_k-releasing lobster and as a result N forms a mental mechanism M_{ik} that goes *on* whenever UCS_k is in the water, causing L_i to flee, we may use A IS AROUND - EPHEMERAL to attribute the content L_k *is around* to M_{ik}'s going *on*. If, unfortunately, on occasion the presence of UCS_k is not caused by L_k, but, *e. g.*, by an ethologist testing our lobster for Individual Recognition, M_{ik}'s going *on* still means L_k *is around*. L_i is misrepresenting the world. This kind of mild misbehaviour we call error, or misrepresentation.

But there is another, more worrying kind of misbehaviour. Consider the following case:

THE PITILESS BIOLOGIST: A biologist experimenting with a group of lobsters releases in the water tank a lobster, L_l, the urine of whom she has somehow rendered odourless. L_l proceeds to bash lobster L_i repeatedly and thoroughly while the biologist releases in the tank a chemical substance UCS_m -which is of the same type as real lobster urine chemical signatures, but has been synthesised in the lab. As a result of this, the token of N in lobster L_i produces a state M_{im} that goes *on* when UCS_m is in the water, and that causes L_i to flee when *on*. Which content should we attribute to M_{im}'s going *on*?

In this case, A IS AROUND - EPHEMERAL is not met: there is no individual such that the homeostatic mechanism of Q^n (the higher order HPC on which N relies) makes it correlate with the cue that has caused N to create M_{im}. A content attribution to M_{im}'s being *on*

32 And if we have an intermediate layer between N and M_{ij} (that is, if the ephemeral contentful state is not a product of the selected-for mechanism, but a product of a product thereof) we will need to have *three* levels of HPCs, and so on and so forth. This means that if, as is probably the case with human mental states, there are many layers of explanations between selection-for and contentful states, the environment has to be stable enough (i.e., there must be homeostatic mechanisms in place) at many different levels.

Figure 9: Two levels of Homeostatic Mechanisms

would be, for all we know, unwarranted. Moreover, we have reasons for refusing to grant content to M_{im}: N continues to exist in contemporary lobsters because it has stumbled upon causal grounds for the correlation of cues of type C with instantiations of properties that make states of the right causal profile fitness-contributing. Our recipes only warrant content attributions when these causal grounds are active in the creation of a certain state. It is reasonable to think that, in case they are inactive, the mental states in question come out contentless[33][34].

3.10.1 *The Internal Perspective*

M_{im}'s being *on* may not have content, but it behaves as if it did. L_i is a one-trick lobster: it smells a chemical signature present in the water during a fight, and then flees (fights) when it smells it again. How come that some of the times the lobster is acting on the content of its mental states and some of the times it is not? The answer is in noting, again, that a content-attribution is but a summary of a certain kind of explanation of the existence of a state and its continuing success. If an explanation of the relevant kind does not exist, it cannot be summarised. And it is completely up to the external world, not the lobster, to provide or fail to provide the necessary ingredients for such explanation.

Still, one (among others) of the things that a content attribution does is providing an explanation of the behaviour of the attributee. And, in this connection, it is true that M_{im}'s being *on* influences its possessor's behaviour in a way that shares certain important traits with the way in which contentful states explain behaviour. We have already identified, in 3.6, the nature of these traits: what M_{im} shares with less catastrophic counterparts is that its positives follow a Procedure of the same kind -because they have been created by the same mechanism. Following the definition presented in PROCEDURE, M_{im}' being *on*'s Procedure is something like: *Go* on *whenever* UCS_m *is in the water, and cause* L_i *to flee.* On the other hand, and as opposed to more succesful products of N, its Procedure does not have an ecologically-fixed context.

I have suggested that abidance by Procedures is the basis of a series of normative claims: in this case we may say that M_{im}, and the lobster that possesses it, cannot be *blamed* for fleeing after smelling UCS_m.

Blameless doings, on the other hand, are only fitness-conducive if the world collaborates. The world needs to provide the properties that make a mental state's causal profile fitness conducive. That is to say, it has to provide for the properties appealed to in clause B2 of EPHEMERAL CONTENTFUL STATE:

B2: There is a higher order HPC Q^n that explains that instantiations of properties C_i go together with the instantiations of the properties in the HPC $Q_i^{n-1} = f(C_i)$, and these lower level HPCs in turn explain the fitness conduciveness alluded to in B1.

There is a second normativity dimension, according to which a mental mechanism does *well* in going *on* only if these properties, *i. e.* the properties instantiations of which are to make the state fitness-conducive, are present:

33 For similar points regarding selected-for states cf. 2.1.

34 I should point out that N, on the other hand, has content: it is a standing proto-belief to the effect that Q^n is in place; something like Q^n *exists*.

NORM: A mental mechanism M_i, produced by a selected-for mechanism N, goes *on* according to its Norm if and only if the properties that the HPC $Q_i^{n-1} = f(C_i)$ keeps frequently coinstantiated with the presence of M_i's Input, and explain its fitness-conduciveness, are present.

That is, for example, a state M_{ij} of lobster L_i, that causes L_i to flee, behaves according to norm if and only if it goes *on* in the presence of a lobster higher than L_i in the dominance hierarchy. M_{im}, the mechanism created as a result of the machinations of the pitiless biology, could very well conform to this Norm.

We may introduce the pair wrongdoing/rightdoing as labels for this second normative dimension:

RIGHTDOING: M's behaviour is right if and only if it accords with M's Norm. Otherwise, it is wrong.

A case of blameless wrongdoing occurs when a mental state, following its procedure, starts to exist in a situation that does not accord to norm. A case of blameful rightdoing happens when a mental state starts to exist according to norm, but without following its procedure. Note, again, that blameful behaviour is not a possibility for these very simple contentful states.

The four combinations of blameful/-less right-/wrongdoing seem to be enough to describe the behaviour of contentless states in a way that makes justice to our internalistic intuitions. So, what are content attributions doing? What are they good for?

Well, Procedures and Norms constrain the behaviour of mental states from opposite directions: Procedures concentrate on the normative dimension of *displaying well-tested behaviour*, while Norms concentrate on *displaying useful behaviour*. I have been defending a view on content according to which contents involve the causal structures that make, in the long run, causal profiles (well-tested behaviour) fitness-conducive (useful). Contents, that is, explain that states that follow its Procedure, reliably accord to Norm:

EXTERNAL/INTERNAL: The content of a mental state, M's being *on*, is a sufficient number of times, the explanation of the reliable connection existing between blameless behaviour and rightdoing by the mental state in question.

To put it simply, contents explain that behaviours are successful -where behaviour is described by the Procedure, and what counts as successful is described by the Norm. Take, for example, content attributions warranted by A IS AROUND -EPHEMERAL: ephemeral mechanisms have the causal profile they have (*i. e.*, they follow the Procedure they follow) because their producer, N, has followed ROUTINE. Now, N's doing so has contributed to its being selected -and, thus, its keeping producing ephemeral mechanisms- because (FITNESS CONDUCIVENESS) the mechanisms it produces have been fitness conducive in their turn (and, if I have defined Norms correctly, this should amount to their behaving according to Norm a sufficient number of times) by relying on individuals -which (HIGHER ORDER HPC) are kept together by a higher order HPC. That is, the complicated state of affairs which underlie content attributions provides an explanation that following

Procedures a sufficient number of times supposes behaving according to Norm.

This reliable connection need not be very good. Only as good as to make M's producer selected for (or to prevent it from being selected against). Contentless states are outside the sufficient number of states for which EXTERNAL/INTERNAL is true. In their case, either there is no reliable connection between blameless behaviour and rightdoings or, if there is, it's explained by bizarre circumstances, not having to do with the higher order HPC that has enabled the selection for M's producer.

A promising idea -although, of course, in need of much further development- is the following: the notions of Procedure and Norm can take care of the behaviour-explanatory features of content attributions in the best way possible that is homogeneous for contentful and contentless states. But a still better explanation of behaviour has to distinguish contentful and contentless states: there is an explanation in the former, but not the latter, of the connection between acting blamelessly and doing the right thing. A complete explanation of a piece of behaviour includes a reason why a mental mechanism with the right causal profile is there to effect the behaviour. This is what intentional explanations of behaviour do.

I would like to hypothesise that EXTERNAL/INTERNAL is true even for sophisticated cognitive creatures such as us. The story would go as follows: judgings using empty concepts such as VULCAN or ILL-STARRED may be blameless -in that there is a Procedure associated with their use- and they may be right -because they conform to some broad Norm such as the ones above- but they will be contentless: there will be no connection between blameless doings and rightdoings. At least no connection of the kind that has explained the past success of the concept-forming procedure[35].

35 I also think, but cannot hope to substantiate here, that the strategy of distinguishing a Procedure-like principle and a Norm-like principle may help solve other debates about normativity. For example, take the debate about the norm of assertion. Williamson (1996 and 2000) dismisses the norms: *assert only what is true* and *assert only what you have warrant to assert*. The main objection to the former is that assertion "obviously has some kind of evidential norm" Williamson (1996, p. 497). The main objection to the latter is provided by situations in which one is warranted to assert something untrue, and assertion seems objectionable. Williamson, instead, defends a knowledge norm: *assert only what you know*. I think this is one of the cases in which norm pluralism is the right option. The warrant norm plays the Procedure role; the truth norm plays the Norm role, and an analogue to EXTERNAL/INTERNAL shows the connection of assertion with knowledge:

EXTERNAL/INTERNAL - ASSERTION: The fact that the content of an assertion *A* is known is, a sufficient number of times, the explanation of the reliable connection existing between blameless asserting and right asserting.

That is, it is knowledge that explains the connection between warrant and truth. This is the actual way in which knowledge is linked to assertion.

4 COMPOSITIONALITY

Let us call *F-involving* those contents that are about F. In my discussion so far I have provided sufficient conditions for the attribution to a mental state (ephemeral or not) of the following F- and a-involving contents, where *F* and G stand for real kinds, and *a* stands for an individual:

- *There is an F around / a is around.*
- *The causal mechanism SHM ensures that when there's an F around there tends to be a G around.*

The goal of this chapter is to canvass the strategy for extending the approach I have been advocating to more complex F-involving contents.

I will, first (4.1) show why the most basic contents must be propositional and not subpropositional. After that (4.2) I show how to provide an interesting set of sufficient conditions for a number of contents beyond *There is an F around S*, such as *There is an F five minutes from here*, *There is an F 500 m away*, etc. The recipe proposed here will simply be an application of what we had already seen in previous chapters -that is, said content attributions will not depend on content attributions to subpropositional components.

Section 4.3 provides a first approach to the task of providing concept-contents (i.e., contents such as *F*) to mental states: I will focus on cognitive setups in which some mental mechanisms (which I will call *collaborative*) contribute to the constitution of a great number of mental states. To avoid problems with attributions of content to states without function, I will concentrate in a philosophical fiction: creatures whose cognitive endowment emerged as a result of mutation and has subsequently been selected for *-swampchildren*.

The first idea to be discussed, in keep with contemporary Causal Role Semantics approaches to this issue, is identifying the concept of F, in one of these interconnected networks of mental mechanisms, with the collaborative mental mechanisms that contribute solely to the constitution of F-involving contents. I will discuss two problems with this approach, which will lead to a (hopefully) better one:

First (section 4.3.1) the problem of cognitive significance. We may have a number of mental mechanisms, such that each of them contributes to the constitution of F-involving contents. Say, as a Hesperus-mechanism and a Phosphorus-mechanism may do with Venus-involving contents. The solution to this problem, it will turn out, involves having a set of mechanisms that create a sufficiently varied pool of F-involving contents. To simplify the discussion I will assume that this set of mechanisms is produced by PRED, a mechanism that effects predications between individual and kind concepts and CONC, a mechanism that creates individual and kind concepts from certain cues.

The need for this set of mechanisms, in turn, leads to a second, more pressing problem (section 4.4): how to fix *as F-involving* the content of thoughts constituted by the tokening of mechanisms whose status as concepts of F depend, precisely, on the content of the thoughts they participate in. That is, if we make the fact that a particular product of

CONC is the concept of F depend on the thoughts it helps constitute, then we cannot make the content of these thoughts depend on the identity of the concepts that form it. In section 4.4.2 I canvass my own positive (and, I hope, non-circular) proposal for fixing the content of concepts, of composition mechanisms and of thoughts. With this I meet the main goal of the chapter. After that, I review (4.5) Millikan's own attempt to resolve the threat of circularity, and I conclude that it is unsuccessful.

As a coda to the chapter, I discuss how my proposal fares compared with other popular accounts of concepts (4.6), I show how the grain of truth in two important insights in the philosophy of mind may be accounted for in the etiosemantic proposal: that association is a pervasive and important mental operation (4.7); and that causal roles may help fix the meaning of concepts (4.8). In particular, it is shown how many of the most vexing problems with Causal Role Semantics may be solved.

All in all, I regard my proposal as building upon the Classical representationalist (Fodorian) insight according to which concepts are terms in a language of thought. One way to see the results of the chapter is as providing an explication of how the atomic components of representations (concepts, and basic ways of composing them, such as predication) acquire their content. This is done while explaining several seemingly independent intuitions in the philosophy of mind, and without incurring in the implausibly radical concept nativism Fodor defends[1].

4.1 PROPOSITIONS FIRST

Most contemporary naturalistic accounts of content are inspired by, and may be considered the sophisticated descendants of, the kind of views about meaning put forward by Kripke and Putnam in the seventies. Those were views about the semantics of names and kind terms, so it is sociologically all too natural that many theorists working in the field see their task as trying to come up with the right account of content for singular and real kind concepts, under the assumption that, once we have those building blocks in place, we can use a version of the compositionality principle to account for propositional contents -where the compositionality principle asserts, roughly, that the meaning of a thought is fixed by the meaning of its subpropositional components plus the way in which they are composed. The literature is rife with examples in which the paradigm of mental representation is the concept of cat (*e. g.*, Loewer (1999), Fodor (1990)) or fly-detectors (*e. g.*, among a great many others Agar (1993) or Zawidzki (2003)). This is unfortunate at least because it makes the following problem daunting:

> What we want is that fly-occasioned "fly"s, and bee-bee occasioned "fly"s, *and representations of flies in thought* all mean FLY. At best, teleological solutions promise to allow us to say this for the first two cases -bee-bee-occasioned tokens are somehow "abNormal" (...) hence their causation

1 And, I hope, without falling in what Godfrey-Smith calls the

> neo-Lockean wild goose chase in search of strangely reified "concepts" or other fundamental representational units. Godfrey-Smith (2009b, p. 38)

that according to him viciates much contemporary philosophy of mind.

> is not relevant to the content of "fly" (...) But teleological theories don't even pretend to deal with the third case; they offer no reason not to suppose that fly-thoughts mean *fly or thought of a frog* given that both flies and thoughts of frogs normally cause fly-thought tokens. Fodor (1990, p. 81)

The problem Fodor is raising in this quote is easiest to see from the perspective of an optimal-conditions causal theory of content such as STAMPE in 1.1.2. Suppose we want to provide the content of the concept FLY according to one of these theories, and we propose the following definition:

> A representation is the concept FLY iff, in optimal conditions, only flies cause it to token.

Even admitting that we have a specification of optimal conditions that rules out little black pellets, mosquitoes, etc. as optimal-conditions causes of the tokening of FLY, there is another kind of causes of instantiation of concepts that simply cannot and should not be ruled out as suboptimal or abnormal: trains of thoughts involving flies may cause further tokenings of the concept FLY, and ought to do so, even if the thinker is temporarily causally isolated from flies. As Fodor goes on to point out, even a perfect thinker, one who never misrepresents, or tokens concepts inappropriately, is such that some of her thoughts are caused by some other of her thoughts.

So, how should we reformulate our optimal-conditions causal theory to filter out not just cases of misrepresentations (black pellets and mosquitoes) but also cases of the concept being caused to token by the tokening of other concepts in a train of thought? Obviously, the quick fix won't do:

> A representation is the concept FLY iff, in optimal conditions, only flies, or appropriately related thoughts, cause it to token.

Leaving aside the extreme vagueness of the condition proposed, thoughts are appropriately related to other thoughts in virtue of their content. Therefore, this cannot be the general strategy for a reductive account of content. This problem can be replicated for any account of content that tries to tie application of a concept to a set of optimal (or Normal) conditions.

Luckily for the work I have been doing so far, this *train of thought* difficulty does not appear in an analysis of *propositional* contents such as *There is a fly around*: it is not the case that, in optimal conditions, other thoughts, in the absence of flies, cause proto-beliefs or proto-judgements with such contents. A perfect thinker would never think *There is a fly around* in the absence of a fly around her. Fodor's problem is a non-issue if one sticks to propositional contents.

But one cannot hope to stick forever to propositional contents. That is, one cannot hope to give an account of sophisticated content-crunching engines such as the human mind without an appeal to proposition-constitutents. I will not review, and will simply endorse, the well-known arguments from the systematicity and productivity of thought to this conclusion -*cf.* Fodor and Pylyshyn (1988), Fodor (2008). Fortunately for the kind of view I am advocating, the compositionality insights stemming from such arguments can be accommodated. My aim in this chapter will be to show how to do this. In the process, I will show

that compositionality, *pace* Fodor -see his (2008, p. 30)-, is compatible with the *propositions first* attitude I have endorsed: one can very well defend that the most basic contentful states are non-analisable representations with propositional content *and* defend that, once a level of subpropositional representations (concepts and the like) is provided for, productivity and systematicity are present in the (now analysable) representations with proposition content.

In the next few sections I will need to use a larger pool of selected-for mental states involving Fs. Before concentrating in thought-constituents, therefore, I will briefly discuss how to provide for content-attributing recipes for a variety of F-involving contents beyond *There is an F around*.

4.2 BEYOND "... IS AROUND".

When discussing the Indeterminacy Problem in chapter 1, I concentrated in the problem of indeterminacy about the real kind the content is about. That is, the problem of finding a reason why a mental state should be attributed with the content *There is an F around* instead of *There is a G around*. This is the most urgent problem, if only because it is the problem that has drawn everybody's attention. But, actually, there is another source for indeterminacy which is seldom discussed in the literature: why should the content be *There is an F around*, rather than *There is an F right in front of me* or *There is an F in a 5 m radius*? I wish to take up now, if briefly, this other question. I have avoided it so far to simplify the discussion, but the resources to solve it are pretty straightforward.

In the Cluster of HPCs, according to the definition I presented in 1.4.5, we have *instantiations* of properties: concrete entities in time and space. I have been making implicit use of some of the properties of these instantions in our descriptions of Indication Profile and Fitness Matrix. For example, if we remember the story DEMOCRITUS AND THE CONTENT OF M'S BEING ON in 1.3, the Indication Profile of M there was recording the probability of M's going *on* [at a time t, and place a], conditional on the instantiation of a property F [at t, near a]. In the Indication Profiles and Fitness Matrices that ground the most basic content attributions, some properties of the event that is the tokening of the contentful mental state may be transformed, via a certain mathematical function, into properties of the instantiation of the property indicated.

With this in mind, a hypothetical way we (or rather, some minor deity) could have used in calculating the Fitness Contribution of Democritus's M is the following:

- We start out with some property F instantiations of which M indicates. In fact, as we have just seen, that M indicates a property or not depends on where and when is such a property instantiated, in relation to when and where does M go *on*. So, more strictly, we start with instantiations of F tagged in space and time[2] in a coordinate system centred in M's position and time when it goes

2 Here again, for simplicity, I'm only considering space and time as the relevant dimension in which M's goings *on* and instantiations of F are to be related. But there could be others. For instance, the relevant dimension may be temperature: M's going *on* at temperature T indicates instantiations of F at temperature T + 10, etc.

on[3]. That is, we rule that M goes on at (0,0,0,0) and calculate the position of the instantation of F in that coordinate system.

- We then need to find out the property instantiations that explain, for each property F such that M indicates it, IP_F and FM_F -see 1.4.1 and 1.4.2. The fact that hits are fitness-increasing also depend on the spatio-temporal location of the instantiation of a number of properties. For example, we said that, for Democritus, $P(on|fly)$ bringing a fitness gain of w_{11}^{fly} depends on $P(nutrient|fly)$ being high.

- Finally, we need to "close the circuit" of property instantiations and workings of M, so that indications issue in positive fitness values. We have:

  1. First activation in M occurs at (0,0,0,0) in a retina-centred coordinate system -that is, it occurs at the place and time of first changes in the retina.
  2. The property F (*Being a black speck*, say) is instantiated at (x,y,z,t) in this coordinate system.
  3. The event in 1 causes an order to the motor-control part of M to move the tongue to position (x',y',z',t')
  4. The property *Being frog nutrient* is instantiated at (x'',y'',z'',t''), very near to (x',y',z',t'). That this tagged instantiation goes together with the one in 2. depends on the homeostatic mechanism of the individual fly -the one that explains that the fly does not disintegrate between its impacting the retina and its being caught by Democritus's tongue[4].

- Once we have closed this circuit we have, for every tagged property F such that M indicates it, a cluster of tagged properties. The cluster, together with the homeostatic mechanism keeping them together throughout selection for M, individuates the natural kind that must figure in the content-attribution. And the spatio-temporal positions of properties in the circuit individuates the location that must substitute the "... is around S". In Democritus's case, actually, we can improve the location in the content-attribution. So, the content should be something like *There is a fly moving from (x,y,z,t) to (x'',y'',z'',t'')*. In the sections to come I will, nevertheless, continue talking of content attributions of *There is an F around*, ignoring this refinement for the sake of simplicity.

3 This is not as straightforward as it sounds. M may be scattered in space and, in Democritus for instance, idealised as it is, it covers from the retina to the motor control in charge of the tongue. The part of M that is relevant for setting the origin of the coordinate system is, maybe, the retina; but our knowing this depends on our having certain insights as regards the causal grounds of the interesting indication-relations: we know that it is changes in the retina that kick off the rest of causal processes we know under the name of M*'s going on*; we also know that the indicated property has a straightforward causal relation with the pattern of changes in the retina, and that such relation depends on the retina more or less facing the place where the indicated property is instantiated, in its proximity.
But, on occasions, the causal relations underpinning the indication-relations may not be so conspicuous. In such cases, looking for the tagged properties that M indicates involves looking also, at the same time, for such causal relations.

4 Incidentally, the fact that we need to rely on characteristics of the individual fly to account for the success of Democritus's mental mechanism may be suggested as a way to distinguish contents involving countable general terms from those uncountable general terms. Contents involving gold, in sufficiently sophisticated cognisers, will need no such reliance on the characteristics of individual chunks of gold.

And now, further content attribution recipes, beyond *There is an F around* are easy to come by. In cases as simple as Democritus it is to be expected that $(x, y, z, t) \simeq (x'', y'', z'', t'')$. That is, the place and time where the fly is first seen and where it is eaten are very approximately one and the same. But this does not need to be so: for example, M may be detecting an early sign of flyhood, say, 30 seconds before the animal appears before Democritus's tongue. Such sign would cause changes in the retina that would, in its turn, cause the protracting of the tongue 30 seconds later.

What should the content attribution be then: *There is a fly...*? It will depend on whether we consider whatever is happening at (x,y,z,t) as *evidence* and whatever is happening at (x",y",z",t") as the *instantiation* of the natural kind, or else (x,y,z,t) as *instantiation* and (x",y",z",t") as a *later consequence*. How to decide exactly *where* and *when* is an HPC instantiated is a complicated question in the metaphysics of real kinds, and one that is outside the scope of this work[5].

It is unlikely that such a mechanism would be very succesful, given how erratic the behaviour of flies is. But suppose that instances of the HPC in question have a predictable enough behaviour, so that part of the property cluster is formed by tagged properties that, first, have a high chance of ocurring "together" and, second, have a time lag between them (this is why the scare quotes in "together") that allows a mental mechanism to detect an HPC some time before it is instantiated. In such cases we may have a mental state, as simple as Democritus's M, with the content, *e. g.*, *There will be a fly in position (0,0,0,+30 seconds) in a tongue-centric coordinate-system*. Many more different contents may be attributed to equally unstructured representations in this way, if the world collaborates.

Equipped with this new lot of basic propositional contents, let me turn now to the discussion of subpropositional ones.

4.3 COLLABORATIVE MECHANISMS

Let us now imagine two creatures, Leucippus and Xenocrates, such that all of their contentful mental states have been selected for -they have no ephemeral ones. Rather, their whole cognitive setup has mutated into existence in some or other of their ancestors. This suggestion is useful, because it allows us to discuss only mental mechanisms with function, and thus avoid the complications we discussed in chapter

5 I suspect that it works more or less like this: instantiations of properties in the Cluster come in "waves", with a few properties showing up first, with low probability; then more and more properties together, with higher probability and then, after a climax of highest density of simultaneous instantiations of properties in the cluster, less and less properties get instantiated with less and less probability. We say that a member of the HPC -a fly, say, or a horse- appears when and where the climax of instantiation density happens. Areas of low probability and low density, before the climax, are evidence; areas of low probability and density after the climax are consequences -and evidence too.
One consequence of this way of putting things is that the distinction between evidence for the presence of F and F itself turns out to be vague. I believe Millikan could take profit of such a vague distinction, to avoid such counterintuitive claims as that, when we hear someone saying that it is raining, we are thereby hearing rain itself -*cf.* (2000, p. 86). What we are hearing, in fact, is evidence for the presence of rain; that is, one of the low density ends of rain itself.
I do not wish to say that evidence for the presence of a real kind is always part of the real kind itself. It is only when the same mechanisms that bring about the most paradigmatic properties in the cluster -in the case of rain: water droplets falling from the sky, mainly- also bring about those low density - low probability tails.

3[6]. Creatures who share this feature with Leucippus and Xenocrates I will dub *swampchildren.* This is, of course, what Fodor would call philosophical zoology: there have never been, and there will never be, an actual swampchild.

Now, Leucippus has four mental mechanisms, A, B, C and D. Their positives have the following contents: *There is a mouse around me now*, *There is a dog around me now*, *There will be a mouse around me in 30 seconds*, *There will be a dog around me in 30 seconds*, respectively[7]. I shall assume that these contents have emerged, independently, as a result of selection for A, B, C and D. I shall also assume that all four mechanisms are independent: the four inputs and four outputs are disjoint. There is no common subsystem going *on* when any of these mechanisms fires.

It is natural to suppose then that, even if there are dogs involved in the contents of both B and D, there is no concept DOG in Leucippus's cognitive setup: there is no constituent that is common to both representations. Likewise for MOUSE.

An alternative setup is Xenocrates's. He also has four mechanisms, 1, 2, 3 and 4, and they also give rise to four contentful states with the same content as Leucippus'. But, instead of each individual mechanism's positives having a different content, the contents are had by states constituted by the joint going *on* of, respectively, (1 and 3), (2 and 3), (1 and 4), and (2 and 4). These mechanisms are wired such that both the mouse being around Xenocrates now and it being around in 30" make 1 fire; both the dog being around him now and in 30" make 2 fire; both the mouse and the dog being around him now make 3 fire; finally, both the mouse and the dog being around in 30" make 4 fire[8]. So, when the mouse is around now, 1 and 3 fire, etc. Besides, the joint activation of 1 and 4 has the same output in Xenocrates than the activation of C in Leucippus; 2 and 4 has the same effect that the activation of D has in Leucippus, etc. See Figure 10 for clarification[9].

There is no competitive edge to Xenocrates' or Leucippus' strategies against each other. Both do the same things in the same situations. Our content-attributing recipes -those developed in chapters 1 to 3 and above- warrant the same attributions to C's going *on* that they do to 1 plus 4 going *on*, etc. But we may still wonder whether states 1, 2, 3 and 4 on their own have content. After all, they have some intuitive claim to being considered subsentential components of Xenocrates's thoughts.

A first suggestion is the following:

6 I will reintroduce such complications in due time. As we will see, we will need the ability to create concepts if we are to credit creatures with the possesion of concepts at all.

7 No ability of entertaining *de se* thoughts is presupposed here. The *me* is simply there to provide an origin for the coordinate-system -see 4.2.

8 This is loose talk. A mouse being here in 30" cannot make a mental mechanism fire *now*. More precisely, what makes 1 fire is the instantiation of a certain property P such that it, together with the input of 4 and the output of these two mechanisms, allow us to fix -in the manner described in chapter 1 and 4.2- a content for the thought constituted of the joint going *on* of 1 and 4 of *There will a mouse around here in 30 seconds*.

9 When reading the figure bear in mind that, in Xenocrates, the presence of any input causes his mental mechanisms to fire, but *all* outputs are needed for the effects to the right to occur. This is what the "AND" and "OR" on top of the arrows mean.
For simplicity, the column of effects in the figure talks about "behaviour appropriate to mice now". This only means behaviour that, together with a specification of the input, and together with whatever relevant features of the environment help fix a content for the positives of the mechanism of *There is a mouse around here now*. All of this I have tried to make clear in chapters 1 to 3.

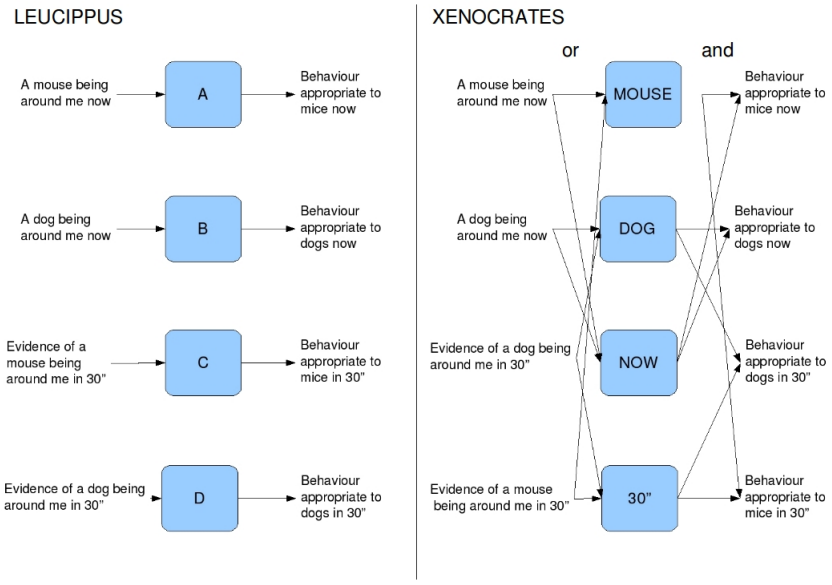

Figure 10: Leucippus vs. Xenocrates

SWAMPCHILD CONCEPT - ALL: A state s of a swampchild is a concept of F if the constitution of every F-involving contentful state involves the instantiation of s, and no other state does.

Under this definition, Xenocrates's 1 being *on* is a MOUSE (swampchild) concept, and his 2 being *on* is a DOG (swampchild) concept. This identification is not entirely implausible. Some evidence: even in this very simple scenario we may see Xenocrates' mechanism 1's being *on* do things that we want tokens of the concept MOUSE to do. For instance, mediating learning about mouses:

Xenocrates can, and Leucippus cannot, recruit another mental state for the input or output of his mechanism 1. Xenocrates may, e. g., recruit a token of the concept WHISKERS (if he had one) for the input of 1[10]. Notice that this does not just mean that judgements of *There are whiskers around me now* cause judgements in Xenocrates of *There is a mouse around me now*. Having WHISKERS and MOUSE connected in this way means that whiskers-thoughts will cause appropriately related mouse-thoughts. For example, if some input causes both WHISKERS to go on and 3 to go on, Xenocrates will form a thought with the content *There are whiskers around me now*, and also *There is a mouse around me now* -in virtue of the causal relation between WHISKERS and MOUSE. But also, *mutatis mutandis*, if some input causes WHISKERS and 4 to go: Xenocrates will form the thought *There will be a mouse around me in 30"*. Contra Fodor -see above-, Xenocrates, simple as he is, is already such that, even if thinking about whiskers causes him to think about mouses, his tokens of MOUSE do not mean anything like *thoughts about whiskers*. In order to have a selected-for concept about thoughts-about-whiskers we would need, first a story that explains the emergence of selected-for *thoughts* about thoughts about whiskers, and Xenocrates's is not such a story[11].

10 This recruitment should be a selected-for mutation. Remember Xenocrates is a swampchild.

11 This is the simplest architecture that solves Fodor's *train of thought* difficulty. But there is no commitment here to the contention that some thoughts (say, about whiskers) need

The account of concepts SWAMPCHILD CONCEPT - ALL although very simple (in fact, we are about to see, *too* simple) is already such that concepts, according to it, are immune to the *train of thought* objection presented in 4.1, and this while being a very straightforwardly teleological account of concepts. It is just not *as* simple as Fodor had in mind. Unfortunately, SWAMPCHILD CONCEPT - ALL as it stands is false. Let us see why.

4.3.1 *Cognitive Significance.*

The problem with SWAMPCHILD CONCEPT - ALL is that there are mental states which have, apparently, similar claims to being the concept of something or other as, say, Xenocrates's 1 but such that they do not comply with the principle. These are the swampchild-analogues of the cases that prompt the conception of concepts as individuated by cognitive significance:

Mental states such as Xenocrates's 1 being *on* have a causal role, described in figure 10. Now, in general we wish to account for the possibility that two concepts of mouse -say MOUSE_1 and MOUSE_2- are such that they have distinct, even disjoint, causal roles. That is, that they be such that each of them participates in the constitution of only *some* of the F-involving thoughts Xenocrates is capable of having.

But SWAMPCHILD CONCEPT - ALL cannot accomodate this possibility. If there are mouse-involving contentful states in the constitution of which MOUSE_1 does not participate (i.e., those in which MOUSE_2 participate), then MOUSE_1 is not a MOUSE concept. This is, I take it, the wrong result. And -as it happens- the easy fix to this problem will not work:

SWAMPCHILD CONCEPT - SOME: A state s of a swampchild is a concept of F if the constitution of *some* F-involving contentful state involves the instantiation of s, and no non-F-involving state does.

It will not work because we wish to allow for the following possibility: as a matter of fact, *all* F-involving contents are G-involving contents, but being an F is not being a G. In such a situation, mental states we wish to count as concepts of F count, according to SWAMPCHILD CONCEPT - SOME, as concepts of G. Suppose, for example, that Xenocrates only had mental mechanisms 1, 3 and 4 -no 2. Then SWAMPCHILD CONCEPT - SOME would yield the result that *all three* mechanisms's being *on* are MOUSE concepts. The role of 4, for example, is to participate in the constitution of some F-involving thoughts (the thought constituted by the joint instantiation of 1 and 4), and does not have as part of its role the participation in any not F-involving thought. But surely 4 is not a MOUSE concept. Only 1, if any, is.

At this point, a possibility is to stick with SWAMPCHILD CONCEPT - ALL, and deny that either MOUSE_1 or MOUSE_2 are concepts of mouse at all. What are they concepts of, then? Maybe of an entity individuated both by mousehood and the concept's causal role. Say, the ordered

always ellicit other, related thoughts (say, about mice). It is overwhelmingly likely that, in more sophisticated cognitive systems, the relations among concepts are not implemented simply in what I have called recruitments, but in more nuanced ways; such that, for example, a thought about whiskers only ellicits the related thought about mice if some other parts of the system carry the information that mice may be relevant to the current plans of the thinker. I have no sympathy for associationism -for more about this, see 4.7.

pair formed by these two things -a *mode of presentation* of mousehood. This, in slight caricature, is in keeping with the Fregean tradition of individuating concepts by their cognitive significance.

One problem with this move is that it would force us to start from scratch: we have provided sets of sufficient conditions for F-involving contents, but the Fs in question have always been some individual or natural kind; it is less obvious how to provide naturalistically acceptable sufficient conditions for mode of presentation-involving contents. And we would need such as set of sufficient conditions if the mode-of-presentation move is to work.

There is another, more important reason why this move is unsatisfactory. One insight behind naturalistic accounts of the mind is that mentality develops as an increasingly sophisticated way of dealing with the world. Under this picture, it is unsatisfactory to conclude that the target of contentful states changes from, say, mouses, to modes of presentation, as we increase the sophistication of our interaction with *mouses* themselves. That is, it is *prima facie* a strange picture one in which cognitive sophistication carries with it a change of the intentional object of thoughts, from mousehood to ordered pairs of mousehood and a certain causal role.

This is not to say that such ordered pairs, or other even more exotic entities, cannot be the intentional object of any thought. On the contrary, in the last few paragraphs I have been discussing these ordered pairs and this (I hope!) will have elicited in the reader thoughts which do have said pairs as their object. This is the right context in which the existence of these contents reveals itself: very sophisticated thoughts such as *The ordered pair of mousehood and a causal role is an unlikely intentional object for Xenocrates's thoughts*. The existence of failures of reidentification caused by differences in cognitive significance, on the other hand, should be dealt with in a different way.

There is another, better way to make SWAMPCHILD CONCEPT - SOME work: having a set of propositional contentful states that is big and varied enough so as to make sure that there will be F-involving contents which are not G-involving for any F and G -or at least that this will be the situation for a majority of concepts:

SWAMPCHILD CONCEPT - BIG A state s of a swampchild is the concept of F if

1. The constitution of *some* F-involving contentful state involves the instantiation of s, and no non-F-involving state does, and
2. The set of F-involving contentful states that the swampchild is capable of having, and involve the instantiation of s, is big and varied enough.

This gets rid of the problem: the version of Xenocrates which lacks mechanism 2 is not capable of having a big and varied enough set of contentful states, so none of his states is a concept according to SWAMPCHILD CONCEPT - BIG. This is, I think, the most reasonable way to credit swampchildren with concepts.

A problem with this solution is the appeal to the size of the pool of concepts being *big enough*. This is a clearly normative condition and, thus, unacceptable for our purposes if left unanalysed -the *enough* is the offending ingredient. In SWAMPCHILD CONCEPT - BIG we are talking of concepts that have been hardwired by evolution; no selected-for producer has created them, and this means that the *big enough* cannot

be unpacked in terms of the conditions for survival and selection of the producer of concepts. That is, there in no "enough*" for this "enough", the way there was, e. g., sufficiently* good indication for Democritus in chapter 1. This is a peculiarity of swampchildren: their states are not the product of other, functional states, but rather were originally mutated from scratch, and thus, *a fortiori*, there are no selected-for conditions on the performance of their producer.

Appeal to an etiological, naturalistic unpacking of *big enough* is possible, instead, if the set of contentful states is the result of the workings of a system, or systems, whose function it is to generate such sets of states, or such that it has as a normal consequence of its functioning the appearance of a sizeable number of collaborative mechanisms à la Xenocrates. That is, *big enough*, and hence the definition of concept, become naturalistically tractable when there is a system selected for the production of contentful states. A corollary to this discussion is that it is unclear that swampchildren may have concepts, even if their contentful states are implemented by a net of collaborative mechanisms, as in Xenocrates' case. Only creatures which have newly emerged contents clearly have concepts.

Thus, in the sequel, I will study the conditions for existence of a couple of mechanisms that are able to endow a creature with a sufficiently varied pool of concepts: A CONC mechanism that creates new concepts, and a PRED mechanism that creates thoughts in which the referent of a kind concept is predicated of the referent of an individual concept. As we will see in due course, once we have these mechanisms in place, it is more reasonable to make the etiology of a certain mechanism M -and not its actual participation in thoughts- what fixes the fact that its being *on* is, or is not, the concept of F. In particular, its having been created by something like CONC.

I turn now to the discussion of these thought-producing mechanisms.

4.4 PRODUCTIVITY AND CIRCULARITY

In previous chapters, we have seen instances of producers that create ephemeral mechanisms with contentful states. We have just been advancing some reasons to think that, without some of these producers, it does not make sense to credit a subject with the possession of concepts. The question now is, what sort of producers should be in place in the cognitive economy of a subject, if we are to have a mechanism s such that its being *on* is the concept of F? An attractive suggestion is that it is enough with a system that has the relational function of creating the concept of F in the presence of a cue C_F, and such that it has created s in presence of this cue. Let us call such a mechanism CONC. We are about to see that spelling out what conditions a mechanism should meet if it is to qualify as a token of CONC are not easy to come by, because of a widespread threat of circularity between mechanism-specification and content-specification.

To start seeing this, consider what is needed to attribute some actual token of CONC with such an etiological function to create concepts. At least, it must be the case that some of its ancestors *have* created concepts in the presence of the relevant cues. How should we adjudicate whether this performance has taken place, that is, whether CONC's ancestors have created *concepts*? Of course, whatever it is that makes those things concepts cannot be that they were created by CONC's

ancestors, on pain of circularity. We have the independent suggestion presented above (now in a formulation suitable for any creature, not just swampchildren):

CONCEPT - BIG: A mechanism s's being *on* is the concept of F if

1. The constitution of *some* F-involving thoughts involves the instantiation of s, and no non-F-involving thought does, and
2. The set of F-involving contentful states that the possessor of s is capable of having, and involve the instantiation of s, is big and varied enough.

And for the appeal to the set of thoughts being big and varied *enough* to be acceptable, as we have seen, we need a mechanism that has the function of producing thoughts, such that its existence depends on the number of thoughts it produces. For simplicity, let us consider a mechanism PRED (from *predication*) that has the function of taking two mechanisms R and S, such that R's being *on* is an individual concept of *a*, and S's being *on* is a concept of F, and outputting a thought with the content *Fa* only if Fa[12].

For a mechanism with this function to emerge, it appears, individual and kind concepts must exist independently. So. at least, it is not straightforward how PRED could help in fixing, in an independent manner, the conditions necessary for CONCEPT - BIG to hold. The remainder of this long section is dedicated to disentangling this circularity, and offering a non-circular account of the nature of CONC and PRED. The following subsection maps some analogies between this problem and the problem of combining the Context and Compositionality principles in philosophy of language. A way to deal with this latter problem will be put to use with the former too. This solution, nevertheless, implies the existence of a core of thoughts whose content is determined independently of CONC and PRED. This may be considered an unwelcome result, so in section4.5 I take some time to discuss an account (Millikan's) that apparently does not have this consequence. Unfortunately, I will conclude that Millikan's way to provide for compositionality and subpropositional thought components cannot work.

4.4.1 *Context and Compositionality*

In Xenocrates, as we have seen, states 1+3, 1+4, 2+3 and 2+4 enjoy a content-endowing explanation of their existence; the theoretical assumption behind the attribution of content to 1, 2, 3 and 4 is that, in virtue of this very fact, there are also content-endowing explanations of the existence of these latter mechanisms. The strategy I have sketched to account for the content of thought-constituents relies on a mental analogue of a version of the context principle:

CONTEXT: A mental state is a concept of F in virtue of the fact that all contents in which it participates[13] are F-involving contents.

12 Together, maybe, with some additional constraints. I do not wish to commit myself to PRED being *the* mechanism that effects predications. It may well be that the mechanism that creates thoughts such as *Mr. Doodles is on the mat* is not the same that creates thoughts such as *Three is prime*. Also, it is possible that predication in human thought has features that are essential to it, and absent from PRED; for example, the *relata* in human predication are assigned with different thematic roles and PRED, we will see, does not do this. By the end of the chapter, anyway, I hope to have substantiated the claim that PRED does effect predication -some, not all, of them; and maybe a simplified, not a human, version.

13 And there is a big a varied enough number of them; see above.

We have not said anything about what makes a content F-involving, apart from enumerating some content schemas (*There is an F around*, etc.) that count as F-involving. But there are a great many other F-involving contents, apart from these. What fixes their content? I have suggested that the answer "being produced by PRED" introduces a problem. The content of the outputs of PRED depend on its inputs in the following way:

COMPOSITIONALITY: The content of a mental state is F-involving in virtue of the fact that a concept of F participates in its constitution.

And, on the face of it, CONTEXT and COMPOSITIONALITY are incompatible. If what makes a content F-involving is the fact that a concept of F participates in its constitution, then what makes something a content of F cannot be its participation in F-involving contents. I cannot go into the complicated debate surrounding the compatibility of top-down and bottom-up determination of content in language and thought. I will only note that one possible strategy for securing such compatibility is not a possibility in the case that interests me. As Szabó points out,

> As long as it is not understood as a causal or explanatory relation determination can be symmetric, so any version of [the principle of compositionality] is compatible with the corresponding version of [the context principle]. Szabó (2008)

But I am interested in what *makes* something a concept of F, and what *makes* some other things F-involving contents. This is precisely the kind of explanatory dependence that, Szabó suggests, may make the two principles incompatible.

In the case of actual learning of the meaning of subsentential expressions there is, though, a simple-minded explanation[14] of how such learning is possible, given the truth of context and compositionality principles. A three-step process is postulated:

1. The meaning of a conveniently big and varied corpus of sentences CS is learned, without making use of the structure of the sentences. That is, the speaker learns to associate each sentence S_i with its meaning M_i, without relying on the meaning of the subsentential components of S_i.

2. The meaning of the subsentential components of sentences in CS is worked out from the meanings learned in 1, with the help of the context principle.

3. The meaning of all other possible subsentential components, and all other possible sentences, is calculated from the information obtained in 1 and 2, with the use of the principles of context and compositionality.

Of course, much more needs to be said about these steps and in particular this last one. The kinds of inferences that allow a speaker to work out the meaning of a new word from the context of utterance and the meaning of the rest of the sentence where the word is found are anything but straightforward. But this is not a problem for the process

14 Suggested to me by Manuel García-Carpintero.

we are interested in: once the function of PRED and CONC is fixed, it is the working of these mechanisms which fixes the meaning of new concepts and thoughts. So, the analogue to step number 3 in thought is pretty clear: the meaning of new concepts is to be fixed by CONC and the cue it uses in the particular occasion of the creation of a concept. *Mutatis mutandis* for thoughts and PRED.

Now, the simplest way to adapt this three-step process to the constitution of F-concepts and F-involving contents is the following Simple Solution[15]:

SIMPLE SOLUTION:

1. The meaning of a conveniently big and varied corpus of mental states CMS is selected for -swampchild like. Each state in such corpus is constituted by the joint instantiation of a number of collaborative mechanisms, but the content of each such a collaborative mechanism plays no role in fixing the content of the states in CMS[16].
2. The meaning of the collaborative mechanisms that constitute the states in CMS is fixed by CONCEPT - BIG. Concepts appear at this stage.
3. The meaning of the rest of possible mental states, not in CMS, is fixed by composing the meanings of collaborative mechanisms, in novel combinations.

It must be noticed that, according to SIMPLE SOLUTION, there are mental states whose meaning depends on the meaning of its constituents -those that appear in step 3. One may worry that this step is always going to be subject to the train of thought difficulty, and others that make a *propositions first* approach sensible -see above, 4.1. I dedicate the following section 4.4.2 to develop a version of SIMPLE SOLUTION that is free from these difficulties. After that, I will take a close look at Millikan's alternative account, that promises to account for productivity without making use (or making, at most, a derivative use) of subpropositional constituents. I will raise a basic point about this account: Millikan's use of mapping functions introduces a covert intentional element and, therefore, is faulty from the naturalistic point of view. The conclusion of the section will be, then, that the appeal to constituents is unavoidable. Hence my positive proposal, which I present now.

4.4.2 *Interlocking Determination*

According to SIMPLE SOLUTION, the way in which we can safely attribute PRED with the ability to take a concept of *F* and a concept of *a* and issuing the thought that *Fa*, is by having a previous stock of concepts *and thoughts*, such that the ancestors of PRED took the former and issued the latter a sufficient number of times, and this explains the existence of PRED. A selective story that complies with this requisite must be of a very particular kind, subject to a number of constraints:

15 I should make clear that I am not transposing a solution to the compatibility of context and compositionality principle from a (more than dubious) assumption that the phylogenetic development of concepts is recapitulated in the ontogenetic generation of individual conceptual repertoires for actual thinkers. I feel no sympathy for such an assumption. Rather, I am drawing this insight from the philosophy of language because I cannot see any other non-circular way in which CONC and PRED may be characterised.

16 That is: collaborative mechanisms do not contribute their meaning to the meaning of the mental state. On the other hand, they *do* have a role in fixing the causal profile of the mental state, which in turn helps fix its meaning.

- The individuation of selected-for thoughts provides one such constraint. Throughout this work I have been making the assumption -typical in teleosemantics- that the fact that selected-for thoughts are copied from one another (i. e., are inherited) helps individuate them: even leaving aside its content, it's not merely its causal dispositions, but its history that makes a mechanism the mechanism it is. Selected-for thoughts, that is, form what Millikan (1984) calls *reproductively-established families*. Under this assumption, if PRED is able to output tokens of selected-for thoughts it must be because these selected-for thoughts have *always* been produced by PRED and its ancestors. Otherwise we would have the problem of explaining on what grounds PRED's output on one particular occasion counts as a token of the type, say, "selected-for thought that *Fa*".

- If there are going to be individual and kind concepts to serve as inputs for PRED, there will have to be collaborative mechanisms such as the ones discussed in 4.3. And these collaborative networks cannot be simply hardwired by natural selection: as we have seen in the previous bullet point, it is PRED and its ancestors that must be effecting the connection of concept into thoughts. Finally,

- While helping selected-for thoughts to recur, PRED must also be securing its own role of taking any individual and kind concepts and issuing a predication thought.

Not many stories (however just-so) are compatible with these constraints. One such story, for the emergence of a mechanism PRED that can produce predicative thoughts[17] is:

INTERLOCKING - PRED A mental mechanism PRED is capable of conjoining concepts in predicative thoughts if

1. A mental mechanism (PRED's oldest ancestor) is mutated into existence. Its causal powers involve, whenever confronted with a certain cue (*e. g.*, the repeated coinstantiation of two mental mechanisms) taking two mental mechanisms (so far contentless; precisely which two mechanisms is a function of the cue) and transforming them in a suitable way (*e. g.*, effecting a conjunctive recruiting). The resulting mental mechanism is also, so far, contentless.
2. For a number of generations PRED outputs a pool of mental mechanisms with significant overlap. That is, PRED is exposed to significantly overlapping cues across generations, in the presence of a pool of significantly overlapping inputs. This grounds an identification of some of PRED's outputs as the same mental mechanisms or states as some of the outputs of PRED's ancestors.
3. Some of the mechanisms that PRED outputs, then, form reproductively-established families in Millikan's sense, and are selected for, say, the indication of a number of properties in a way which (following the content-attributing recipes I

17 Again: maybe only a subset of predication-thoughts, and maybe only a kind of predication-thoughts that are distinctly non-human in that they do not assign thematic roles and the like.

have presented elsewhere in this work) allow univocal content attributions to them. Only from this moment on, these recurring outputs of PRED count as selected-for *thoughts*.

4. After a sufficient number of these thoughts have come into existence, CONCEPT - BIG allows us to attribute thought-constituent content to some of the inputs to PRED -cf. 4.4. Only from this moment on, these recurring inputs to PRED count as selected-for *concepts*. This provides a foothold for evaluating the relation between PRED's cue and its inputs and output (which are now contentful) in "human-readable", fully intentional terms. This is why we need parallel selection for PRED and the thoughts it produces.
5. PRED has helped produce a number of selected-for thoughts. If it is now to have the ability to produce ephemeral thoughts of the same kind it must be because the kind of HPC hierarchy I reviewed in 3.3 and 3.8 is also present here:
   a) Each selected-for thought produced by PRED involves the presence of one or more HPCs.
   b) Consider a number of pairs of a cue S_{Fa} and the selected-for thought that PRED creates in its presence, *Fa*. There must be a higher order HPC linking cues such as S_{Fa} with thoughts such as Fa in a sufficient* number of cases.
   c) In such a situation, if PRED goes on to associate an individual concept *b* with a kind concept *H* to form an ephemeral mental state in the presence of cue S_{Hb}, the output mental state may be attributed with the content *Hb* -the state of affairs that the higher order HPC keeps together with the cue.

In this way we are assigning PRED a non-negligible task in the production of thoughts, while still allowing for independent content determination of both its outputs and its inputs. After a number of generations of this interlocking usefulness of PRED and the states it creates, PRED is ready to compose new, not selected-for states into meaningful thoughts.

Once a PRED mechanism is in place, the formulation of sufficient conditions for a mechanism that outputs new concepts is relatively straightforward. I am envisaging a mechanism CONC which, in the presence of a cue S_a, produces a concept of a. Again, there needs to be an interlocking story for this mechanism to help create concepts that have an univocal content:

INTERLOCKING - CONC: A mental mechanism CONC is capable of creating concepts if

1. A mental mechanism (CONC's oldest ancestor) is mutated into existence. Its causal powers involve: whenever confronted with cues of a certain type, producing a mental mechanism whose positives are, so far, contentless.
2. For a number of generations CONC outputs a pool of mental mechanisms with significant overlap. This is because it is exposed to significantly overlapping cues across generations. Some of these recurring mechanisms, then, form reproductively-established families in Millikan's sense.
3. PRED takes some of these recurring mechanisms and forms selected-for associations with other mechanisms -created

by CONC or emerged independently- that can, thereby, be attributed with content following the usual recipes in chapter 1 and the beginning of this chapter.

4. We can now use the content of the thought created by PRED to fix the content of these products of CONC which start being concepts. From now on, the relation between CONC's cue and its output may be evaluated in intentional terms. Finally,
5. CONC has helped produce a number of selected-for concepts. If it is now to have the ability to produce ephemeral concepts of the same kind it must be because the right HPC hierarchy is present here:
   a) Each selected-for concept produced by CONC is about an HPC -*i. e.*, every thought created by PRED with a mechanism produced by CONC involves an HPC, and every thought produced using a certain mechanism produced by CONC involves (maybe among others) the same HPC.
   b) Consider a number of pairs of a cue S_a and the selected-for concept that CONC creates in its presence, A. There must be a higher order HPC linking cues such as S_a with HPCs such as A in a sufficient* number of cases.
   c) In such a situation, if CONC goes on to create an ephemeral mechanism B in the presence of cue S_b, B may be understood as the concept of b -the HPC that the higher order HPC alluded to in (b) keeps together with the cue.

Finally, we can use CONC to give a characterisation of concepthood:

CONCEPT: A mechanism M is a concept of *a* if

1. Cue S_a has caused CONC to create M, and
2. a is the HPC that the higher order HPC alluded to in INTERLOCKING - CONC 5b connects with S_a.

I should like to note that what endows CONC with the ability of creating concepts -that is, contentful mechanisms- is not simply the fact that it has the relational proper function to create mechanisms that satisfy CONCEPT - BIG, such that this function is satisfied in Normal situations by a particular set of conditions. This kind of Millikanian individuation conditions will not do for the same reasons why we found them wanting in chapter 1: the notion of Normal conditions is tied to the conditions that were operative during selection for the functional device in question, and there are bizarre (but, as far as we can tell, nomologically possible) scenarios in which CONC acquires that very relational proper function without there being any HPC, or any other causally-grounded correlation to facilitate the emergence of said function. The consequence is that there is no HPC, in turn, to fix the content of the associations among mechanisms created by PRED -again, for exactly the same reasons that gave rise to indeterminacy in our early examples of frogs and flies. This is a crucial obstacle to the whole project of providing a content for the outputs of CONC.

The etiosemantic position on this issue is to insist that CONC's content-endowing capabilities do not depend solely on its having the right kind of function with a set of Normal conditions -as in teleosemantics- but, rather, that they depend on there being the right set of normal conditions (in this case, a nested set of HPCs) that enables the emergence of

the right kind of function for CONC. Functions are very important, but second to HPCs.

We may note now that many of the states created by PRED and CONC may lie outside the domain of the HPCs appealed to in INTERLOCKING - PRED and INTERLOCKING - CONC, but now, having provided for subpropositional constituents and predication, we have the possibility of modelling the content of these thoughts with the help of, say, interpreted syntactic trees. We have, that is, all of the advantages of the traditional picture of compositionality while still retaining a theory that is naturalistic through and through -and that has the advantages of the *propositions first* approach.

Before going on to draw some conclusions from this sketch of a theory of contents, I discuss Millikan's alternative, involving what she calls mapping functions.

4.5 MILLIKAN ON PRODUCTIVITY

We may distinguish two important kinds of productivity which, *prima facie*, appear to be very different. One goes by the name of *indexicality*: there is a sense in which, whenever I think *There is a food here now* (I will sometimes call this *the food thought*), I am thinking a wholly new content -at least because it has never been "now" before: that there is food there and then. Each token of *There is food here now* is relevantly different to other tokens of the same thought: each happens at a different time; and each means a different proposition: *There is food at <the place in which the thought is tokened, the time at which the thought is tokened>*. A cogniser who is able to entertain the food thought already shows a limited but very real kind of productivity: she is able to think indefinitely many contents, and contents never entertained before by anyone, just by tokening the food thought at different times.

But most would take this to be productivity in, at most, a honorific sense. What they have in mind, rather, is *compositionality*. A thought system is productive in this other sense, roughly, if it counts with a vocabulary of concepts and of modes of composition, such that the content of a thought is determined by the content of the concepts that compose it and the way in which they are composed. The productivity allowed by compositionality is way richer than mere indexical productivity. For example, someone who possesses n individual concepts (such as *Michael*, or *Eve*) and m kind concepts (such as *horse* or *shoemaker*) and the operation of predication will be able to entertain n times m different contents. And, one is tempted to say, *really* different, not just indexicality-different.

Indexicality is compatible with a *propositions first* approach -see 4.1: the productivity afforded by indexicality does not depend on the recombinability of a vocabulary, or the iterability of a number of syntactic structures but on features of the propositional thought such as the time or place in which it was tokened. On the other hand, compositionality provides for much richer productivity but, as we have seen, it leaves open the problem of how to provide a naturalistic account of subpropositional contents in light of, for example, the *train of thought* difficulty. Millikan's answer to this conundrum is interesting and original: she sets out to show that, appearances notwithstanding, indexicality and compositionality are, at bottom, two aspects of the same phenomenon. According to this picture, if indexicality can be explained in a frame-

work which takes propositional thoughts to be the basic bearers of content, compositionality can be as well.

In Millikan's theory, productivity in intentional systems is but one manifestation of the articulated character of natural signs in general -*cf.* Millikan (2004, p. 48). This character, perhaps in a concealed way, is already present in the most simple relations of indication. Take the definition which kicked off our discussion:

INDICATION: A mechanism M's going *on* indicates instantiations of a property F around S iff

I1: $P(F|on) > P(F)$ and

I2: The difference in probabilities in I1 is causally grounded.

Notice that, according to the definition, the relation of indication holds between events *such that some of their features are a function of one another*. Namely, their spatio-temporal location: instantiations of F <*near* M, *now*> are indicated by M's going on <*where* M *is, shortly after*>.

This kind of causally grounded relations between event types may be described using what Millikan calls *mapping functions*: mathematical transformations from features of the sign to features of the indicated states of affairs. In the case just discussed, M's going *on* at (x, t) indicates an instantiation of F at $f(x, t)$. The relevant mapping function here is MF_M, where $MF_M(x, t) = (\text{near } x, \text{shortly before } t)$. More terminology: this mapping function, in turn, individuates what Millikan (2004) calls a *system of signs*: the class of possible and actual signs obtained by varying the relevant parameters. In the case just described, the relation of indication established between M and F individuates a system of signs, SYS[18], that may be characterised as follows:

$$\forall x, t\,(\text{M's being } \textit{on} \text{ at } \langle x, t\rangle \in SYS)$$

For concreteness, suppose that F is the property *Being food*, and M's being *on* is a natural sign of this property because the presence of a round, orange thing nearby always causes Ms to turn *on* (and is the only thing that causes them to turn *on*), and most round orange things are peaches (and hence food) in M's surroundings.

Under these assumptions, the causal grounds appealed to in clause 2 of INDICATION only support some of the mappings from signs in *SYS* to states of affairs. That is, while there are many members of *SYS* (*i. e.*, instances of M's turning *on* at a certain time and place) that have been or will be caused by the presence of a round, orange thing that is food, there are very many other members of SYS for which it will not be the case that the corresponding state of affairs (the presence of food there and then) will also occur, or even made more probable by the causal mechanisms in place. Some clear cases are

- M's being *on* at $\langle \text{around here}, \text{the distant future}\rangle$,
- M's being *on* at $\langle \text{Mars}, \text{now}\rangle$

at least if we assume that peaches will not be the predominant orange round things in the distante future around here, or now somewhere in Mars, and there are is no other abundant, edible, orange and round

18 Notice that M belongs to *SYS* only relative to the natural-sign relation it establishes with F. It may be that M establishes other natural-sign relations with other properties, each of which will define an alternative system of signs.

stuff. More interestingly, the causal grounds of the indication relation will also fail to support the mapping function for many everyday values of *x* and *t* -*e. g.*, those that pick out events of M's turning *on* that are not caused by the presence of peaches.

Natural signs as described by INDICATION are a plausible precursor of indexicality. Take, again, the food thought: *There is food here now*. This thought is indexical because the content it expresses depends on the time and place at which it is tokened. Now consider M's being *on* (for all we have said, M's being *on* may or may not be the same thing as the food thought). It is a natural sign of the event consisting of the instantiation of food somewhere sometime, and we have cashed this out as saying that events of M's being *on* belong in a system of signs, *SYS*, such that each member of *SYS* has, according to MF_M, an image in a range of (possible or actual) events of instantiations of food. The role that the thought-type *the food thought* plays in the traditional description of indexicality is played here by the system of signs *SYS*.

As we are about to see, Millikan's main idea is to recognise the existence of increasingly complicated mapping functions for increasingly abstract systems of signs; one of the very complicated, very abstract examples will amount to what is traditionally understood as compositionality. But, first, we need to see the role of mapping functions in the attribution of contents according to Millikan. The idea is that contentful states (*intentional icons* in her terminology) are related to the state of affairs that is their content (roughly, their *real value*, in her terminology), thus:

> When an indicative intentional icon has a real value, it is related to that real value as follows:
>
> 1. The real value is a Normal condition for performance for the icon's direct proper functions.
> 2. There are operations upon or transformations (in the mathematical sense) of the icon that correspond one-to-one to operations upon or transformations of the real value such that
> 3. Any transform of the icon resulting from one of these operations has as a Normal condition for proper performance the corresponding transform of the real value. Millikan (1984, p. 99)

For our current purposes, claim 1 plays the role of the recipes for the content of different contentful states I have been giving throughout this work. But we are currently interested in claims 2 and 3: 2 can be paraphrased, roughly, as saying that, whenever a state has content, the relation of that state to its content is covered by a mapping function; while 3 says that the rest of states in the same system of signs have as content their image according to this mapping function. In this way we can provide, for example, a content attribution for the state consisting of M's being *on* a year from now: namely, that there will food near that token of M in a year. This is so because MF_M takes the former (merely possible) tokening of M to the latter (merely possible) instantiation of food.

In my example in this section I have been considering a mapping function, MF_M, that transforms the spatio-temporal location of the sign into that of the signified event -*cf.* Millikan (2004). We may now

note that mapping functions may take just about any feature of the sign to any of the signified event. For example, in some domains -say, a Mediterranean beach- there is a mapping from the *distance between footprints* on the sand (sign) to the speed at which some hiker was going (signified). Or, in some other domain -say, a field near Toulouse- you may find a mapping between the *number of apples* fallen on the ground and the *speed of the wind* during the past few hours. INDICATION can be generalised so as to include these other cases:

INDICATION GENERALISED: A sign of type s's having feature F_S indicates an event of type E with feature $f(F_S)$ iff

I1:

$$P\Big(\text{An event E with feature } f(F_S) \mid \text{A sign of type s's having feature } F_S\Big) > P(\text{An event E with feature } f(F_S))$$

and

I2: The difference in probabilities in I1 is causally grounded.

Where, as I have said, the mapping function f may transform whatever features of sign and signified -not just spatio-temporal locations. Any such mapping functions may be fed into steps 2 and 3 in Millikan's quote.

Thus, if, *e. g.* bee dances -Millikan (1984, p. 107)- are such that transformation of some of their features (number of loops, angle of the axis of the eight, etc.) correspond to transformations of features of the position of the nectar, actual bee-dances share a system of signs with bee-dances-after-transformations-of-features, and these latter entities have as content their image according to the mapping function that helps individuate the system of signs -see above. For example, if an actual dance D has as content that nectar is 50 m from the hive in the direction of the sun, the fact that dances are members of a system of signs determined by the mapping function that takes dances to positions of nectar has as a consequence that a hypothetical dance D* in which the waggle part is a hundred thousand times longer than in D has, as content, that there is nectar 5000 km from the hive in the direction of the sun. Give or take.

The Naturalistic Worry.

Taking stock, what we have seen so far is a plausible description of what a natural sign is, encapsulated in INDICATION GENERALISED, and how such a picture may help explain the limited kind of productivity we call *indexicality*: certain mapping functions take features of the sign (and, by extension, of simple contentful states) to features of the signified event (and, by extension, of the content of simple contentful states). A mild version of the naturalistic worry that I will advance against the application of mapping functions to compositionality also afflicts its application to this simpler indexical case. I will present the worry now; and, after discussing Millikan's approach to compositionality, I will show how to extend the worry to the more interesting case.

The relation of a natural sign to its signified, according to Millikan's picture as summarised above, may be described at three different levels:

L1 The first level is constituted by the *concrete causally-grounded relations* that are established by signifier and signified. That is, *e. g.*, the very causal relations that tokens of M that have actually existed established with the presence of food. Relations such as: a certain peach's being around a token of M at a certain time causing M to activate[19].

L2 Then comes the level of the *causally-grounded indication relation* between types of events: every pair of signifier and signified that is covered by the causal underpinnings of the concrete causally-grounded relations in L1. For example, only a class of distances, neither too large nor too small, indicate that a hiker was walking at a certain speed[20], and this has to do with causal (particularly, physiological) constraints enforced by the muscular and skeletal arrangement of the human body. These constraints fix the class of pairs of signified and signifier that is covered by the indication relation.

L3 The causally-grounded indication relations in L2 may be only probabilistic, and most of them will only be effective in a small, gappy domain. So, finally, we may wish to abstract mathematical transformations that fill-in and extend the domain in which the indication relation holds. These are the *mapping functions* -the $f(x)$ in INDICATION GENERALISED. For example that to each possible distance between footsteps *dbf*, corresponds a speed of the hiker $s(dbf)$, or to each possible waggle dance D a position of nectar $n(D)$.

All three levels are needed for the Millikanian picture summarised above to work. Level L2 -that of the causally-grounded processes which cover the concrete pairs of sign-signified- provides signified events for as yet uninstantiated signs. This is what we need for a productive system of signs, which was the whole purpose of the exercise. Level L3 -constituted by the mapping functions themselves-, in its turn, is needed at least if we want to provide signified events for members of a system of signs that lie beyond the causally-grounded domain, such as the aberrant waggle dance D*. Apart from these abnormal cases, there may be other, more everyday examples in which an appeal to level L3 is needed. Say, a token of M inside the skull of a creature that has been abducted from its original habitat and placed inside a cage in a lab -where the causal explanation of the correlation of orange-and-roundhood with nutritiousness is entirely different from that in the wilderness.

If level L3 does real work in content attributions[21], it is reasonable to worry about the naturalistic credentials of the resulting theory of

19 And causing, or maybe constituting the fact, that food be there

20 For clarification, let me show how this example is a substitution of the INDICATION GENERALISED schema:

- The sign s is a set of footsteps.
- The relevant feature F_S is the distance between footsteps in s.
- The signed event E is the hiker's walk.
- Finally, her speed is the relevant feature of the event, $f(F_S)$.

21 At the end of this section I briefly discuss whether Millikan is committed to level L3 or not.

content. The problem is that there are no facts in the causal order to determine that the content of D* is fixed by the mapping function that yields 5000 km as a result, and not another function that yields any other value -or, maybe, another that has gaps for values not covered by the causal underpinnings of the relation between dances and nectar position. In choosing one of these mapping functions as *the right* one, then, we are surreptitiously introducing an intentional element in the theory.

Millikan has some things to say about the closely related issue of Kripkestenian worries about our ability to follow rules, but what she has to say does not solve this naturalistic worry:

In her (1993, chapter 11), Millikan discusses the problem of rule-following as introduced in Kripke (1982). Kripke issues a sceptical challenge against theories of meaning that make facts about the dispositions to use linguistic expressions on the part of speakers constitutive of the meaning of these expressions. Kripke puts forward two different arguments -*cf.* also Boghossian (1989, p. 509). In summary, they are as follows:

First, the *infinite truths* argument: there are infinite truths about the use of some expressions; for example, there are infinite true substitutions of the schema *a plus b is c*. But -even if we leave aside our dispositions to make mistakes, cf. Kripke (1982, p. 26f)- our dispositions are finite, being the dispositions of finite beings in a finite amount of time. So, it cannot be that these infinite truths are accounted for simply by relying on our dispositions.

Second, the *normativity* argument. There are facts about the correct way in which we should apply our terms. That is, a theory of meaning should account for the fact that terms *ought* to be applied in some ways but not in others. Now, there is no way to read an *ought* off a disposition. Dispositions can only tell us how things are, not how they should be.

Although Boghossian (1989, p. 528) defends that causal-informational theories are, for the purposes of the sceptical argument, a subset of dispositional theories of meaning, it is not clear that he was considering teleosemantic theories among the former. In any event it seems that teleosemantics has some resources to answer both Kripkean worries. Millikan (1993, p. 217)'s strategy is to argue that purposes to conform to unexpressed rules are *biological purposes*. The idea, as the reader has probably anticipated, is to place biological functions at the base of the normativity of meaning. The *ought* of meaning is a biological *ought*, which can be subsequently unpacked in naturalistically unobjectionable terms by an etiological theory of functions such as, say, Millikan's own theory of proper functions -*cf.* Millikan (1984, chapter 2f), Millikan (2002). The infinite truths of the first objection, on the other hand, flow naturally from these normative facts: facts, *e. g.*, about what ought the terms to apply to cover an infinite number of cases.

Millikan's example involves the mating strategy of male hoverflies. She identifies a "proximal hoverfly rule". If the male is to intercept a female in flight,

> the male must make a turn that is 180 degrees away from the target minus about 1/10 of the vector angular velocity (measured in degrees per second) of the target's image across his retina. Millikan (1993, p. 218)

This, plausibly, is not simply a disposition that male hoverflies have, but, rather,

> the hoverfly has within him a genetically determined mechanism of a kind that historically proliferated in part *because* it was responsible for producing conformity to the proximal hoverfly rule, hence for getting male and female hoverflies together. Millikan (1993, p. 219)

This kind of historical properties of the mechanism warrant our attribution to it of a biological function -or, in this context, a biological purpose. If this is correct, Millikan can then give an answer to Kripkensteinian sceptical complaints:

- Infinite truths: the hoverfly mechanism has the function of, given the angular velocity of a retinal shadow, issuing a muscular command that makes its possessor fly in a particular direction. This is so for an infinite number of angular velocities or, at any rate, for a number that far surpasses the number of actual uses that actual hoverflies will make of the mechanism.
- Normativity: the biological function of the mechanism, attribution of which is warranted by the kind of history that it has, underwrites the relevant normativity claims made as regards its functioning. Intrinsically, mechanisms *ought* to comply with their function.

It is clear that this teleosemantic response goes *some* way towards answering the sceptical challenge, and, this, at least, warrants a closer examination of the theory -beyond Boghossian's somewhat unfairly-lumped category of "causal-informational theories". What I wish to discuss now is the *scope* of the teleosemantic solution. Given that it is the hoverfly mechanism's causal history that supports the attribution of biological purposes of it, it is natural to consider that features of the history may constrain the scope over which the biological purpose is operative. In this case, the selection for the hoverfly mechanism has occurred because a couple of indication relations are in place:

I1:

$$\begin{aligned} &P\Big(\text{A female hoverfly being at x,t}\big| \\ &\text{angular velocity of retinal image being } \omega\Big) \quad > \\ &P\left(\text{A female hoverfly being at x,t}\right) \end{aligned}$$

I2:

$$\begin{aligned} &P\left(\text{Intercepting female hov. at x,t}\middle|\text{Displaying behaviour B}\right) \quad > \\ &P\left(\text{Intercepting female hov. at x,t}\right) \end{aligned}$$

where $\omega = f(x,t)$ and $B = g(x,t)$. That is, the relevant indication relations hold under

1. Certain transformations of angular velocities of retinal images onto positions of female hoverflies, and
2. Certain transformations of behavioural responses onto positions of female hoverflies.

Millikan's "proximal hoverfly rule" may be rendered thus:

PHR: In presence of a retinal image with angular velocity ω, issue behavioural response $B = g(x,t) = g \circ f^{-1}(\omega)$[22].

Now, what is causally grounding I1 and I2? Well, the average flight speed of hoverflies remains approximately constant, because hoverfly physiognomy remains approximately constant; non-hoverfly darting things are sufficiently sparse, and it remains this way because, among other things, the ratio of non-overfly insects vs overflies is also approximately constant, etc. This kind of facts make it the case that the inequalities I1 and I2 hold. But, crucially, *only insofar as said causal grounds do ground the indication relations*.

Let us suppose that these causal grounds are operative only for values of ω below 330 degrees per second -I am making this up-; the problem should now be apparent: there are infinitely many mathematical functions that overlap with f in the range supported by the causal grounds, and infinitely many others that overlap with g. And there is absolutely nothing to determine which one of them should figure in PHR.

It should be noticed that a number of things Millikan says against some alternative candidates for PHR have no bearing against the present worry: suppose that never in the history of hoverflyhood a female has produced an image in the retina with an angular velocity between 500 and 510 degrees per second. It is still the case that the following "proximal *quoverfly* rule" is wrong Millikan (1993, p. 221):

PQR: In presence of a retinal image with angular velocity ω, issue behavioural response $B = g^*(x,t) = g^* \circ f^{-1}(\omega)$.

Where

$$\begin{cases} g^*(x,t) = \textit{Don't move} & \textit{if } 500 < f(x,t) < 510 \\ g^*(x,t) = g(x,t) & \text{otherwise} \end{cases}$$

Hoverflies do not have the biological purpose of following PQR: it is not *that* rule that explains that males catch females. There is a principled reason to choose PHR over PQR: there is a concrete causal explanation of the fact that the behaviour of male hoverflies is fitness-conducive. This explanation involves the causal underpinnings of the relations I1 and I2, and these causal grounds also cover the range of angular velocities between 500 and 510 degrees per second, regardless of whether such values have or have not been actually instantiated.

There is another, more complicated case that Millikan considers: suppose that, because of engineering constraints, hoverflies do have a blind spot between 500 and 510 degrees per second. So, their dispositions are best described with PQR. As a matter of fact, whenever a shadow between that range of velocities crosses a male's retina, it doesn't move. Millikan (1993, p. 222) claims that, in this case, the rule the male hoverfly has the biological purpose to follow is still PHR: the disposition to rest at ease in the blind spot in no way furthers the hoverfly reproductive goals[23]. In the way I have been putting things, the causal grounds tying

22 Where $g \circ f(x) = g(f(x))$.

23 This is, I think, the sensible position. At the end of this section I will discuss Millikan's apparent change of mind in this respect.

retinal shadows with future positions of female hoverflies are operative also in the blind spot; on these grounds we should include those values in the rule[24].

But all of this still gives no reason to choose one among the many different functions that overlap perfectly inside the zone of causal grounding and diverge, however wildly, outside of it. That is, Millikan has given no reason to decide among the different substitutions of the following proximal hoverfly rule schema:

PHR-SCHEMA: In presence of a retinal image with angular velocity ω, issue behavioural response $B = g_i \circ f_i^{-1}(\omega)$.

where $\forall i \left(g_i \circ f_i^{-1}(\omega) = g \circ f^{-1}(\omega) \right)$ inside the causally-grounded domain of the function.

Millikan wants PHR to come out as the one and only rule male hoverflies follow, but the kind of considerations she advances -having to do with what rule explains the fitness-conduciveness of the hoverfly mechanism- cannot in fact distinguish PHR from an infinite number of competitors -the infinitely many substitutions of PHR-SCHEMA. Another way to put this point is the following: mathematical functions such as *f* and *g* have a role to play in the causal explanation of the selection of a mechanism only insofar as they describe the behaviour of whatever it is that is causally effective in said selection. But causal mechanisms[25] underdetermine which mathematical functions describe them. This underdetermination leads directly to rule-indetermination[26].

Notice that it will not do to retort that *the* mapping function has a set of normal conditions for application (that yields the causally-grounded domain) and that, outside of this set, the right thing to say is that the application is abnormal. In fact, the foregoing discussion has shown that there is no fact of the matter as regards which is the right mapping function outside of the causally-grounded domain[27]. So, finally, this provides reasons to remain appropriately circunspect in our appeal to mapping functions. Mapping functions, I submit, are well and good if we restrict their application to the causally-grounded domain: we should build our content theory only upon the relations recorded in

24 There is a certain complication I am putting aside here. As it stands, the case is underdescribed: the causal underpinnings of 12 depend, among other things, on the mean velocity of male hoverflies. If the engineering constraints alluded to in the description of the case are such that the maintenance of this mean velocity depends on leaving this blind spot in the response to retinal shadows, then this is a true gap in the causal underpinnings, and, *pace* Millikan, there is no principled reason to include these values in the rule. Another possibility is that engineering constraints do not mess up with the causal grounds for the indication relations in this or other ways. If so, we can endorse PHR. This second option is the one I'm taking for granted in the main text.

25 At least of the kind that do not have universal application, *i. e.*, those whose workings cannot be embedded under strict physical laws.

26 This discussion should not be taken to mean that I endorse the conclusion of Kripke's sceptic. My point is simply that *Millikan's theory* leaves, at least, the indeterminacy described by PHR-SCHEMA.

27 Millikan, in personal communication, has suggested that appeals to the needs of the consumer (the male hoverfly) can do more to fix the content of the biological purpose of the hoverfly, and thus the particular function that must go in PHR-SCHEMA, that I am according here. The male needs a female hoverfly, so that is what the biological purpose is about.
I am not sure about that. The needs of the consumer can, surely, decide among different purposes *within the range in which the consumer will use such purposes*. But, *e. g.*, reacting to extremely high or extremely low angular velocities of retinal shadows would never be conducive to fulfilling the needs of the consumer, because such velocities will never indicate the presence of a female hoverfly.
Appeals to the consumer leave open a fair amount of indeterminacy among mapping functions.

INDICATION GENERALISED. So, for example, a sensible teleosemantics should admit that D* (the bee dance with an aberrantly long waggle part) is meaningless.

In a recent discussion, Millikan appears to agree with this conclusion (*beemese* is the name Millikan gives to the mapping function that takes bee dances to positions of nectar):

> It is unlikely that a dance that, by logical extension of beemese rules, would tell of nectar much too far to fly to could be either danced or, more central, recognized by fellow bees. No ancestor bees have had dispositions to make use of such dances. Such bee dances, then, are meaningless in beemese. Millikan (2006, p. 107)

It is informative to see in which way Millikan's diagnosis of the situation differs from the one I have been offering here. On the one hand, Millikan relies on the empirical implausibility of dances such as D*: maybe bees are unable to dance them. Maybe so, but, in the discussion of Kripke's sceptic I have been reviewing, Millikan has strived at separating content from actual dispositions. It may well be that no bee has ever had the disposition to use D*, but in the parallel discussion, the fact that a hoverfly had a blind spot between 500 and 510 degrees -and, thus, had no dispositions to respond in that range- was -correctly, I think- dismissed as irrelevant for the purposes of content attributions. If so, it is difficult to see why a lack of disposition to respond to D* should matter. Either dispositions are irrelevant or they are not, but Millikan cannot have it both ways. Besides, what happens if, after all, bees are able to dance the dance? Suppose that the mechanism that creates dances has a tendency to create, very rarely, an aberrant dance such as D*. If I am right, we are still forced to say that D* is meaningless: the causal grounds that cover the relation of typical dances to positions of nectar do not cover D*, and, thus, there is no fact of the matter regarding which mapping function should we apply to it. But now it is unclear what Millikan would want to say about this case, and on which grounds.

On the other hand, Millikan talks of the impossibility of such an aberrant dance being recognised by other bees. *Recognition* is an intentional notion: presumably, that there is recognition depends on whether the receiving bee is able to form a mental state with the same content as the dance. We have no idea whether this is possible or not, and we should not care: there is no need, for a dance to have content, that such contentful mental states exist. The dance may well be issuing orders directly to the muscles of the bee without the intervention of the bee's cognitive system -though in point of fact dances do not, of course. That there is recognition is not necessary to fix the content of dances.

Millikan reaches the right conclusion -that aberrant bee-dances are meaningless- but by, first, making the content of dances depend on mapping functions and, then, restricting the scope of these mapping functions to those supported by actual dispositions of the consumers of the representation. This goes against the grain of her proposal regarding Kripke's sceptic and, in fact, makes it essentially a dispositional account of the kind that were the main target of Kripke's discussion. The right way to restrict mapping functions is, I have claimed, by attending to the causal grounds of these very mapping functions -the natural sign relations of INDICATION GENERALISED.

Millikan is happy (even if maybe for the wrong reasons) to accept that some bee dances are meaningless. But she is not willing to accept an analogous result in the case of human thought. Undoubtably we are able to think about events which are causally isolated from us, and Millikan wishes to honour this tenet. The next paragraph casts a doubt on the resources of her theory to do so: it is even more unclear that mapping functions are able to fix the content of thoughts than they are to fix the content of bee dances.

Compositionality

As I advanced in section 4.5, Millikan's ultimate goal is to make both indexicality and compositionality particular cases of the general productivity afforded by mapping functions. We are now in a position to see how may one think of compositionality as depending of mapping functions of the same kind as the ones that accounted for indexicality.

Remember from above that a simple contentful state such as [M's being *on* here now] -which, I said, means *There is food here now*- belongs in a system of signs, *SYS*, together with all other actual or possible events of M's being *on*. Members of *SYS* and their signified events are tied together by a certain mapping function, and I have just been arguing that we have a grip on this function only within the causally-grounded domain.

On the face of it, compositionality is an entirely different beast: productivity is achieved by the more or less free recombination of conceptual items into more or less iterable syntactic structures. There does not seem to be any clear place for mapping functions from thoughts to propositions in this story. Millikan makes the interesting proposal that there actually *is* a causally-grounded mapping function from beliefs to states of affairs[28], just like from bee dances to positions of nectar.

The system of signs here is, roughly, the class of all possible beliefs. In simple systems of signs such as *SYS* above, you could get from one sign to another by modifying their spatio-temporal location. In the belief system of signs, the "feature" that must be modified to get from one sign to another is more elusive: the main transformation is *substitution*, an operation that taks, say, the thought *Democritus jumps* to, on the one hand, thought like *Xenocrates jumps* and, on the other hand, to thoughts like *Democritus protracts its tongue*. Likewise, the state of affairs consisting of Democritus's jumping transforms to the state of affairs consisting of Democritus's protracting its tongue, and to the state of affairs consisting of Xenocrates's jumping. The set of possible sentences reachable by transforming a thought s defines the ways in which the state of affairs s represents should be considered as articulated. A state of affairs plus a certain way of articulating it individuates what Millikan calls a *world affair*[29].

Let us call the mapping function which takes the system of signs which is the class of every belief to their meanings $\mathrm{MF}_{\mathrm{Mentalese}}$. How

28 *World affairs*, really. See below.

29 Millikan is after a fine-grained notion of state of affairs, according to which "Theatetus swims" and "Theatetus exemplifies swimming" are different states of affairs because they are differently articulated. If transformations define articulations, it may be suggested that states of affairs are articulated in *every* way. For example, there is a straightforward transformation that takes "Theatetus swims" to "Theatetus exemplifies swimming" -substitution of predicates.
In response to this, Millikan may, perhaps, defend that there are ways to distinguish relevant from irrelevant transformations. In any event, I do not wish to press this point any further.

are we to establish that the belief-system maps onto meanings according to $\mathrm{MF}_{\mathrm{Mentalese}}$? Bear in mind that this mapping function must suffice to endow with meanings beliefs that have never been entertained before by anyone -the whole point of introducing mapping functions, after all, was to account for productivity. I will not worry about how to account for something similar to Evans (1982)'s Generality Principle, according to which, if a thinker is able to entertain the thought *Fido is brown* and the thought *Bill Gates is tech savvy*, she will be able to entertain *Fido is tech savvy*. It is very difficult to see just what in the causal order is going to make $\mathrm{MF}_{\mathrm{Mentalese}}$ take *Fido is tech savvy* to the proposition that Fido is tech savvy, but it is also open to Millikan to defend that we cannot really think this thought. There does not seem to be any straightforward way to adjudicate this issue.

It is best to concentrate in an uncontroversial subset of $\mathrm{MF}_{\mathrm{Mentalese}}$'s domain. Consider again the food thought, *There is food here now*, as entertained by a human thinker, and all other thoughts that derive from the food thought by substituting *here* and *now* with other spatio-temporal concepts, say, *inside the Pinatubo volcano*, or *three million years into the future*. It is clear that we can think that there is food at these places and times, and, if Millikan's account of this ability is correct, this is because $\mathrm{MF}_{\mathrm{Mentalese}}$ takes, *e. g.*, the thought *There was food inside the Pinatubo volcano during the 1991 eruption* to the proposition that there is food then and there.

Now, there are certain causal connections between thoughts of the food-thought kind and facts having to do with the location of food: food's being somewhere sometime has caused the tokening of certain thoughts which, in their turn, have caused fitness-improving (say, food-grabbing) behaviours. These causal facts may help ground the part of $\mathrm{MF}_{\mathrm{Mentalese}}$ that makes reference to places and times in the domain that humans occupy -even if the particular place and time has never been and will never be occupied by a human being- but they cannot ground thoughts that make reference to location outside this domain, for exactly the same reasons that I presenteda above, when discussing the naturalistic worry. It is only that, in the human case, we cannot simply bite the bullet and say that the thought *There was food inside the Pinatubo volcano during the 1991 eruption* is meaningless. That this thought means what it seems to mean is nonnegotiable.

Millikan has suggested (1984, 2004, 2006) that a mechanism that tests beliefs for inner consistency may help explain our coming to have beliefs about world affairs which are causally isolated from us (let us call them *far away beliefs*), and which in no way further our biological goals:

> Consistent agreement in judgments is evidence that ... various methods of making the same judgment are all converging on the same distal affair, bouncing off the same target, as it were. If the same belief is confirmed by sight, by touch, by hearing, by testimony, by various inductions one has made, and is confirmed also by theoretical considerations (inference is a method of identification too), this is sterling evidence for the univocity of the various methods one has used to identify each of the various facets of the world that the belief concerns. Millikan (2006, p. 111)

So, let us suppose that I am told that water boils at 100ºC outside my light cone[30], and independent theoretical reasoning lets me reach the same conclusion. Here, according to Millikan, the consistency in these two judgements works as a confirmation of the relevant hypothesis about water. Even if it's true that the workings of a consistency tester would be enough to fix a mapping function that deals with far away or useless beliefs[31], the problem with off-causal-grounds mapping functions turns into this other problem: there is no fact of the matter as to whether a certain mechanism is a *consistency* tester. To see this, consider what it takes for a certain mechanism, let us call it CONSIST, to acquire the function to test a corpus of beliefs for consistency. At the very least, for CONSIST to acquire such a function, there must be beliefs such that it is a CONSIST-independent fact of the matter whether they are consistent or not. Otherwise, if all there is to two beliefs being consistent is that a token of CONSIST gives a positive output when confronted with them, the relation of being consistent is entirely vacuous.

Let us assume, then, that there are consistent beliefs prior to the existence of CONSIST. We may want to hypothesise the following three step process:

1. Beliefs in a certain corpus CB acquire their meaning (and their status as consistent or inconsistent with one another) independently of CONSIST.

2. CONSIST tests beliefs in CB for consistency, and thereby acquires the function of being a consistency-tester.

3. CONSIST helps fix the meaning of other beliefs by testing for their consistency with beliefs in CB and previously tested beliefs.

The problem with such a story is that the description of 2. is tendentious. It is unwarranted to claim that CONSIST is testing beliefs for consistency, where *consistency* is a relation that holds between any beliefs whatsoever, far away or not, useful or not. The only matter of fact is about the following: CONSIST tests beliefs in CB for consistency*, where consistency* is consistency between beliefs about states of affairs in causal contact with human beings -these are the kind of beliefs the content of which will have been fixed in step 1. And a consistency* tester clearly cannot help fix the meaning of beliefs about far away beliefs.

The upshot of this discussion is that mapping functions cannot play the role Millikan wants them to in the explanation of the compositionality of beliefs -not even with the help of consistency* testers. It is wrong to think of a mapping function between beliefs and world affairs such as $\mathrm{MF}_{\mathrm{Mentalese}}$ as a *precondition* for beliefs to acquire meaning. The idea that there is *the right mapping function* to play this role is already fully invested with the intentionality we are seeking to explain.

Let me summarise: in Millikan's account of productivity, the main bearer of mental content is the (propositional) thought. Although she recognises a sense in which concepts such as DOG have meaning, this

30 By the way, whatever happens outside the light cone seems to be the stock example in discussions of teleosemantics and the reference to far away, causally-isolated facts -*cf.* Peacocke (1992). I should like to note that many events outside our light cone are really near from us: the events going on a metre away from me five Planck times in the future -I am aware that this is not a rigorous way of talking- are outside my light cone, but it is clear that teleosemantic accounts may be able to deal with thoughts involving that spatio-temporal location -which is just around the corner, and almost now. Anyway, let us assume that we are dealing with water boiling *well outside* my light cone.

31 See Rupert (1999) for some reasons why it may not be.

meaning is entirely dependent on the thoughts in which the concept participates in the following way: such thoughts have their meaning provided by a mapping function, and it is invariances in the world affairs to which all DOG-involving thoughts are taken by such a function that fix the meaning of DOG. All of these world affairs involve dogs, and it is in virtue of this fact that DOG refers to dogs. Unfortunately, as we have seen, the appeal to mapping functions is unable to do the job it was hoped to do.

We need to stick to the classical idea according to which there are thoughts whose content is *determined* by their structure and the content of their subpropositional components. We need, that is, *bona fide* compositionality. Thoughts such as *Bill Gates is tech savvy* are composed by the concepts BILL GATES and TECH SAVVY, and the operation of predication; and the meaning of *Bill Gates is tech savvy* derives (via a compositionality principle) from the meaning of its constituents and the way in which they are organised. The sensible content-naturalising program, I submit, involves providing an account of the meanings of concepts and of mechanisms able of performing syntactic operations among them. We may then simply model the meaning of thoughts with abstract structures -say, interpreted syntactic trees. This is entirely compatible with recognising the existence of other dimensions of meaning, such as Millikanian *real value* (the world affair to which thoughts correspond according to causally-grounded mapping functions, if there is one) and *sense* (the causally-grounded mapping function itself). We should abandon the hope of accounting for productivity simply by using mapping functions, if we remain committed to naturalism.

The alternative I recommend, of course, is the kind of interlocking determination I described above. I will now finish the chapter drawing some consequences of my view, and comparing it with other popular accounts.

4.6 OTHER THEORIES OF CONCEPTS

The interlocking-determination picture helps clarify why the appeal of several popular theories of concepts, while explaining what they get wrong[32]:

4.6.1 *The Classical Theory of Concepts*

In the so-called *Classical Theory* of concepts[33], a concept such as BACHELOR is learned by sticking together the concepts UNMARRIED and MALE which are already part of the conceptual repertoire of the learner, and which, together, provide necessary and sufficient conditions for the presence of bachelorhood. Such more basic concepts will also have their definitions in terms of other, even more basic concepts until, eventually, the whole conceptual system will bottom out in undefined primitives (in the Modern Empiricists's original version of the theory, such primitives would be perceptions, experiences or some such).

Concepts, according to the Classical Theory, have necessary and sufficient conditions of application. If to be a bachelor is to be an unmarried male, the application of BACHELOR requires the application

32 In the rest of this section I'm drawing from my Master thesis.

33 In this review I'm drawing from Machery (2009), Laurence and Margolis (1999) and the papers in Margolis and Laurence (1999).

of the concepts that express the latter properties, and just them. In this way, the Classical Theory offers a very elegant account of concept acquisition. But the Classical Theory is wrong, for a number of reasons. First, apart from maybe BACHELOR and a few others, not many concepts have plausible necessary and sufficient conditions of application. This is mirrored in the fact that not many words can be, clearly and without reminder, defined. Apart from the empirical difficulty of finding such definitions, there is also Quine (1953)'s well-known attack on the notion of analiticity which can equally be considered an attack on the idea that concepts have definitions at all.

Available empirical data also militate against Classical Theory: belonging to the extension of a Classical concept is an all-or-nothing question. If an entity has all the necessary and sufficient properties to belong, it is in. Else, it is out. There are no other cases, and there are no subcases. Therefore, such account of concepts does not explain typicality effects. The fact that we are quicker in identifying typical instantiations of a concept, such as nightingales for BIRD or fox-terriers for PET goes unexplained as far as the Classical Theory is concerned. Besides, the attempts to demonstrate experimentally the psychological reality of definitions have been unsuccessful -*cf.* Fodor et al. (1999).

But, maybe, the most crucial reasons for the failure of the Classical Theory are the problems of ignorance and error, as described by Kripke and Putnam. In order to apply the concept GOLD we don't need to have the concept of, say, ATOMIC NUMBER, or any other that provides necessary and sufficient conditions of application. We can be way more ignorant than that, and only have vague ideas about gold being a very expensive, typically yellow metal. This is compatible with our being proud owners of the concept of GOLD. The problem of error is similar: we can be wrong about some of the characteristics of the entities that our concepts refer to without failing to have the concept. For instance, we may firmly believe that philosophers are, as such, irresistibly inclined to pessimism. But, even if as a matter of fact there are optimistic philosophers, this does not mean that we don't have the concept PHILOSOPHER. These two problems of ignorance and error are absolutely omnipresent in the field of real kind concepts: precisely, such concepts are characterised by our ability to refer to this or that substance, about whose underlying properties we may be ignorant, or hopelessly wrong.

How does the theory I have been sketching account for the intuition that concepts such as BACHELOR are intimately linked to other concepts such as MARRIED or MAN? It is perfectly possible that a mechanism such as CONC works in the following way: whenever it is presented with properties of several kinds, it creates a new concept that follows a Procedure (see chapter 3.6) according to which it is to fire if these properties are present. For example, CONC creates a concept of H and issues an order to PRED (see section 4.4.2) with the content that, whenever something is both F and G, it must create a thought with the content that it is H. CONC's having done like this in the past has been fitness-contributing: this is, maybe, because it has stumbled upon a causally-grounded correlation between properties of the same type as F and G, and properties of the same type as H, such that whenever something has the former, it is sufficiently* likely that it will have the latter.

In such a setup, a thinker may have the disposition to think that something is H if she also thinks that it is F and G. But the content of the concept H is not *F and G*; it is, rather, that substance that correlates with F and G thanks to the same causal grounds that have made CONC fitness-contributing overall. Of course, this is only one of a great number of ways in which a concept-producer such as CONC may have made itself useful. It may use all sorts of other cues for the formation of concepts, and make them behave according to all sorts of alternative Procedures. The very general characterisation of concept that we have provided allows for that. This, in particular, explains that many concepts are not like BACHELOR, in that their application is not tied to the application of any other concepts -put in more familiar terms, in that they lack clear definitions.

CONC may make concepts behave according to more sophisticated procedures. For example, it may make PRED compare every perceived individual with a prototype for bachelorhood and, if the overlap is sufficient, raise the possibility of a positive veredict -such that the slightest additional hint that the individual in question is, say, unmarried, precipitates a formation of the thought that he is a bachelor. This is one of innumerable ways in which the general framework I defend may explain typicality effects. That this mechanism, or any other, is actually present in our cognitive systems is irrelevant for the purposes at hand[34]; no particular mechanism -no particular set of cues and Procedures- is tied to the nature of concepts.

As regards the problem of error: there is no clear limit to the quality of the correlation between the properties that belong in the cue for application of the concept and the property reference of the concept itself. The correlation between *unmarried man* and *bachelor* is, to be sure, particularly good, but correlations with the much lower quality of the one between *pessimistic* and *philosopher* may have been instrumental in making a certain concept-producer such as CONC fitness-contributing. If such a set of Procedures make PRED form useful thoughts in a sufficient* number of occasions, and there are sufficient causal-grounds for this, CONC may have obtained its concept-producer credentials even if its concepts follow Procedures which are not particularly good.

4.6.2 *The Prototype Theory*

The Classical Theory of concepts mirrors old-school descriptivism in semantics: with a great deal of simplification, this is the thesis that proper names and natural kind terms are synonymous with descriptions that uniquely pick out their referent. When the problem of lack of appropriate definitions was already well known, Searle (1958) tried to solve it by appeal to the idea of a cluster of definitions. Proper names, according to Searle, would refer to whatever it is that satisfies a "sufficient but so far unspecified number" Searle (1958, p. 171) of said descriptions. This solves several of the problems of old-school descriptivism: at least, it explains the fact that Manuel Sacristán, *e. g.*, could have failed to be a philosopher without thereby failing to be Manuel Sacristán; this would be so because the candidate to be Manuel Sacristán would still satisfy a sufficient number of other descriptors.

A theory of concepts that can be understood as partially exploiting Searle's insight is the Prototype Theory. According to it, concepts de-

34 Although, of course, *some* such mechanism must be present.

ployment is mediated by a statistical evaluation of the most prominent features of the property that the concept expresses. So, for instance, the concept PHILOSOPHER endows the features ENGAGES IN SPECULATIVE DEBATES or WRITES PHILOSOPHICAL PAPERS with a high weight, the features PESSIMISTIC and TEACHES PHILOSOPHY with a middle-to-low weight and WHITE BEARDED or ABSENT-MINDED LOOKS with low weights. Someone will fall under the extension of the concept PHILOSOPHER if the weighted average of her features goes above a particular threshold value. Prototypical instances of the concept (such as Manuel Sacristán) will have particularly high weighed averages; this is why it will be easier to identify Manuel Sacristán as a philosopher than, say, a philosophically-informed professional surfer. One problem with Prototype Theory is that of missing prototypes: uninstantiated (such as MEDIEVAL COMPUTER) or heterogeneous kinds (such as MACHINE) do not typically have statistically relevant properties. Another is the problem of compositionality: the prototype for WOODEN SPOON (a typically large spoon used for cooking) does not arise readily from the prototype for SPOON (a small metallic spoon, most probably) and that for WOODEN OBJECT (if there is one). A famous example by Fodor is PET FISH: a prototype fish is, maybe, something like a cod; a prototype pet is most probably some kind of dog or cat; but the prototype pet fish is a goldfish.

The general form of the explanation why this theory is appealing is already familiar: a number of concepts are such that the Procedure that is used to deploy them in some contexts involves such statistical weighing of properties, or something similar. But these Procedures do not fix the reference; they are an effect of the workings of a concept-producer such as CONC, and what really fixes the reference of a concept is the entity that corresponds to the cue CONC has used to create this concept, according to the causal grounds that have made CONC fitness-contributing.

Under this picture there is no temptation to think that prototypes should compose. The Procedure used to deploy the PET FISH concept may have a convoluted relation with the Procedures used to deploy PET and FISH; nothing in the theory makes it the case that such Procedures should simply compose. There may be, for example, a first pass in which something is deemed to be a pet fish if it is deemed to be both a pet and a fish, but there may also be a pet fish - prototype on top of that to speed up the process if and when it is possible. There is no limit to the sophistication of the Procedure for this concept; but its reference will remain unaffected because it depends only on matters having to do with its producer and the cue the latter used to create the former.

4.6.3 *The Theory-Theory*

Theory-theorists suggest that to have a concept is to have the folk equivalent of a scientific theory. To have the concept NUMBER is to know how it meshes with other concepts, what inferences are we entitled to make by using the concept, etc. A concept is individuated by its role in one of these folk theories, just like in Kuhnian philosophy of science a theoretical term is individuated by its role in the theory whence it belongs. The theory-theory makes an effort to accommodate Kripkean essentialist insights. When people try to categorise a particular entity under a concept or another they can have, in their folk theories, what

has been called an essence placeholder -*cf.* Laurence and Margolis (1999). People can accept that something can have all of the habitual superficial features of gold while failing to be gold, because people's GOLD-theory allows for the possibility of there being the wrong thing, whatever it may be, occupying the slot of the essence placeholder. But this proposal won't do. Reserving a place for an essence placeholder in a theory is like appending the conjunct "... and has an internal essence" to a Classical description. Such placeholder and such conjunct are totally unspecified, mirroring the almost null knowledge that the lay person has about the essence of her real kind concepts. But, in fact, golds's essence is highly specific, having to do with a certain electronic structure -that of elements of atomic number 79. The GOLD-theory and the SILVER-theory may have, thus, the same essence placeholder (for a suitably uninformed lay person, or a kid), while their essences are very different. Some people's theories of BIRCH and BEECH, the famous Putnamian example, may be exactly the same ("it's a tree", maybe), and an essence placeholder won't tell them apart: both concepts will have it. This is just Millikan's warning:

> There has been the tendency in the psychological literature to misinterpret Kripke's and Putnam's antidescriptionist views on the meaning of proper names and natural kind terms as invoking definite descriptions at one level removed. (No, Kripke did not claim that the referent of a proper name N is fixed in the user's mind by the description "whoever was originally baptized as N," nor did Putnam claim that the extent of a natural kind term is fixed for laymen by the description "whatever natural kind the experts have in mind when they use term T". Millikan (1998)

I will have something more to say about the identification of concepts with positions in an causal network in section 4.8 below. Let me say here though that, indeed, how a concept meshes with other concepts helps fix the content of concepts. This happens in step 5 of INTERLOCKING - CONC: after CONC has produced several recurring mechanisms which then PRED associates with other mechanisms, something like CONCEPT - BIG fixes the meaning of the products of CONC attending, precisely, to their causal role in the network that results in selected-for thoughts. This is the non-negligible contribution of causal roles in the fixation of concept-meaning. But, once CONC has emerged with the right content-endowing history, virtually any concept can have virtually any role. BIRCH and BEECH are a case in point. What fixes their meaning is CONC together with the cue it used to form them in each case -most likely, in this case, a word encountered when reading a book. This cue goes together with a kind of tree, according to the causal grounds that have made CONC fitness-contributing, as discussed above.

4.6.4 *Frege Puzzles*

The very simple cognitive architecture consisting of CONC, PRED and the thoughts and concepts they form is already powerful enough to explain the appearance of Frege puzzles: CONC creates concepts in the presence of certain cues, and these concepts have as reference the individuals or kinds that a certain higher order HPC connect with the cues in question. Now, there is no guarantee that every cue will

correspond to a different entity -that every cue that prompts CONC to create a concept will correspond to a different referent. To be sure, in a majority* of occasions different cues *will* correspond to different referents -if CONC has the kind of history sketched in INTERLOCKING - CONC- but sometimes they will not.

When they don't, there will be two different concepts with the same referent. Besides, each such concept will have a Procedure of application, which will be a function of the cue that has prompted CONC to create it. The difference in Procedure explains that one concept of the pairs be used to form some beliefs and the other, some others. This is what difference in cognitive significance boils down to. So, against the neo-rationalist contention that, essentially,

> Concepts C and D are distinct if it is possible rationally to judge some content containing C without judging the corresponding content containing D. (Peacocke 2008, p. 60)

We may offer the following alternative:

CONCEPT IDENTITY Concepts C and D are distinct if it is necessary to appeal to two distinct acts of creation by CONC to account for their existence.

That is, concepts are distinct if CONC has created each one independently. This, in turn, explains that Peacocke's principle of cognitive significance provides excellent evidence of the distinctness of concepts. The point is, simply, that differences in cognitive significance are not what make two concepts distinct, but a consequence of their being distinct.

This proposal about the identity of concepts is not strictly referentialist either, if Fodor is right about referentialism:

> The question at issue is which, if any, of the beliefs, desires, etc. in which a concept is engaged is constitutive of its identity. Referentialists say: 'None of them; all that matters is the extension'. Fodor (2008, p. 87)

What I am suggesting is that a concept is individuated by an act of creation by a concept-producer. Facts about its content (which, indeed, is exhausted by its reference), and about "the galaxy of beliefs, desires, hopes, despairs, whatever, in which the concepts are engaged" Fodor (2008) (which are simply facts about the concept Procedure, in the very complicated case of human cognisers) flow naturally from facts about the circumstances of its creation. CONCEPT IDENTITY offers a nice way of tying together these features of concepts that Fodor takes, implausibly, to be independent.

4.7 THE TRUE RELEVANCE OF ASSOCIATIVE MECHANISMS

It is maybe not totally implausible that PRED, the mechanism of predication in thought, has been directly selected for -by creating states that have also been selected for. But for many other mechanisms a different story must be true: they must have been created by selected-for mechanisms, or, in any event, by chains of mechanisms that have a selected-for mechanism at the end. It is a challenge to give truth conditions for the last link in one of these chains, given that only the first is selected for.

In the case of PRED -let us suppose it has not been directly selected for- what we would need is a story that fixes a certain much-higher-order HPC that has as a particular case the connection between S_{Fa} and Fa. That is, we need a higher-order mechanism, HOM, that, on the basis of certain environmental cues, creates, on one occasion, a mechanism that effects predications -that is, PRED- and, on another, a mechanism that does something else. This is apparently difficult to obtain, in the face of what I take to be a sensible requisite about the HPC that facilitates selection for HOM: it must comply with what we could call the *Homogeneity Constraint*.

HOMOGENEITY CONSTRAINT: The inputs of a selected-for mechanism must be homogeneous.

It is not easy to make precise what this homogeneity must amount to, but examples are easily given: we have been tacitly following this principle in the selected-for mechanisms we have described in the past chapters:

- The mechanism N that created individual-involving mental states in lobsters had, as its input, *any combination of a urine chemical signature and an outcome in fight*. That is, all inputs had to be homogeneous in that they were all members of this kind.
- The mechanism that creates individual dances in a bee has, as its input, the position of a source of nectar, relative to the hive. All inputs have to be of this kind.

The Homogeneity Constraint stems from the fact that HOM was *selected for*. Such selection happens because mechanisms perpetuate themselves through reproduction; that is, offsprings are relevantly similar to their ancestors, and this includes their having relevantly similar causal powers: roughly the same kind of things makes them react. If offspring-mental-mechanisms reacted to things of a radically different kind from their ancestor-mental-mechanisms, a doubt whether they are member of the same reproductively-established family -and, thus, a doubt whether we were witnessing a process of selection of one *kind* of mechanism- would be very much in order.

Now, HOMOGENEITY CONSTRAINT seems to go hand in hand with the fact that the mechanisms a higher-order mechanism creates have states with contents which are also homogeneous among them:

- For N, mechanisms whose being *on* have the content *Lobster #i is around.*
- For the bee-dance mechanism, *There is nectar <there>*.

This is what makes a mechanism HOM that has created PRED and other, different cognitive mechanisms unlikely. What may these other mechanisms be, that are relevantly similar to PRED, in the same way that contents such as *Lobster #i is around*, for different #i, are similar? This is, I think, a genuine problem, and one that has passed largely unnoticed by teleosemanticists.

There is a solution if a homogeneous cue type gives access to a variety of HPCs -that is, if cues of the same type connect with, on one occasion, the relation of predication and, in another occasion, very different metaphysical relations among properties. As I have suggested

in chapter 3, one such cue may be the frequent cofiring of mental mechanisms. This homogeneous cue may give access to a large varieties of homeostatic mechanisms, because it is relying on the following, extremely general HPC:

BARE-BONES CAUSAL GROUNDS: The HPC such that

1. One of its seeds is formed by:
   a) A pattern of frequent cofiring of two mental mechanisms A and B.
   b) The causal grounds $CG_{A,B}$ of this frequent cofiring.
2. Its specialised homeostatic mechanism is the causal grounds of the fact that (a) and (b) go normally together.

The grounds of the frequent coinstantiation between (a) and (b) that 2 is talking about is, simply, the following: without a causal ground such as $CG_{A,B}$, the rate of cofiring of these two states depends on chance, and *chance in general is less efficient than causation in sustaining coincidence*. This seems like a very basic fact about the world, and one available everywhere and everytime.

A number of higher-order mechanisms may rely on the BARE-BONES CAUSAL GROUNDS HPC (from now on, also BBCG) and thereby get selected. Long term potentiation is, probably, one such mechanism (see 3.4), but there are surely others. We may hipothesise that something similar to BBCG is the most basic HPC relied on in the evolution of the brain, and the conferral of content to its states. BBCG gives access to very different causal structures from a set of homogeneous cues: there may be very little in common to the reasons that may make two mental mechanisms cofire regularly.

Part of the past appeal of associationism -the contention that thought is a chain of mental associations- probably stems from a realisation that frequently co-ocurring mental states, being evidence for the presence of causal structures in the external world, may have been a most basic raw material in the emergence of mentality[35].

4.8 THE TRUE RELEVANCE OF CAUSAL ROLES FOR SEMANTICS.

The main insight behind Causal Role Semantics[36] (CRS) is that

> the semantic properties of a mental representation are partially constituted by certain causal or inferential relations between that and other mental representations. Loewer (1999, p. 120)

The account of concepts and compositionality in thought sketched in this chapter may help explain the appeal of CRS: causal roles do indeed

35 If something like this is correct, Ryder's SINBAD neurosemantics (2004, 2006) may come out as a special case of etiosemantics. Ryder defends that representation occurs in SINBAD networks (*i. e.*, pyramidal cells in Sets of INteracting BAckpropagating Dendrites), which "have a powerful tendency to structure themselves isomorphically with regularities in their environment" (2004, p. 212). Although I will not discuss the proposal, I will say that the neural-computational details of Ryder's proposal are fascinating, and that it looks quite possible that something like this be true of some of the actual contentful mental states in our brains.
On the other hand, the issues with compositionality we have been dealing with in this chapter are left uncommented by the theory as it stands. SINBAD neurosemantics is, so far, a theory about atomic representations. Besides, it cannot explain the possibility of representation in systems without pyramidal cells -simpler or alien brains, for instance.

36 Also Conceptual and Inferential Role Semantics. I will not distinguish between the three.

help fix the content of mental representations -via what I have called *collaborative mechanisms*, *cf.* 4.3. It is needed that PRED -a mechanism that effects predications- creates a big and varied enough pool of F-involving thoughts for there to be a selected-for concept of F, which will help in its turn fix the concept-creating role of CONC. So, in the early stages of the development of PRED and CONC, the mental configurations in which the state that will later be the concept of F participates play a crucial role. These mental configurations are nothing but causal roles.

The relation between causal roles and contents, though, is not straightforwardly one of constitution. Something is a concept of F, I have said, if it has been created by CONC in presence of the right kind of cue. For each individual concept, then, there is no restriction on the number of thoughts in which it is involved. That is, no restriction on which causal role does it have.

Even so, it is true that causal roles help fix the content of a certain concept C. But it is the causal role of *other* products of (ancestors of) the producer of C -the ones that have endowed CONC with the ability to create concepts, through a story such as INTERLOCKING - CONC. As a final argument in favour of my view, I would like to show how the account of concepts presented here deals with the most common objections to CRS[37]. This, together with the fact that it explains why and how causal roles are relevant for the semantics of concepts, maybe helps see it as a plausible alternative to CRS proper.

4.8.1 *Error*

Actual causal roles involve dispositions to err. Very few speakers (more likely none at all) are such that they have the disposition to always apply CAT correctly. A common reaction from CRS theorists is resorting to ideal causal roles. But it remains to be seen how such appeals to ideality may be made consistent with naturalism.

Broadly teleosemantic accounts of content such as my own have as a main goal the treatment of error. In the case of simple contentful states, such as the ones discussed in chapter 1, the strategy was not identifying the content of a state with whatever it happens to indicate, but with the HPC that enabled the emergence of a state with a function to indicate -cf. chapter 1 for details. The answer to the Error Problem for CRS is a complication of this other answer. We may have the disposition to judge that Fa in certain situations in which, in fact, $\neg Fa$. But what fixes the meaning of that thought is not the actual dispositions of the judger, but the interplay of:

1. The workings of PRED (the mechanism that effects predications in thought), together with S_{Fa} (the cue that has prompted it to effect a predication in this case).

2. The workings of the producer of concepts CONC, together with S_F in one case and together with S_a in the other.

Those functions are fixed, in part, by the causal role of their products in earlier generations: whatever it was that explained that effecting a state with such-and-such as causal role in reaction to thus-and-so a cue is the meaning-contributing factor. The relation between this factor and causal roles is complicated enough to allow plenty of room for

37 In the exposition of the objections I follow Block (1998).

error, and there is absolutely no guarantee that the cue used for the production of the thought that Fa necessitates that Fa.

4.8.2 *Holding together the two factors in two-factor CRS*

Block (1986, 1998) defends a *two-factor CRS*. According to this version of the theory, there are two dimensions of the meaning of a thought T:

- One entirely dependent upon T's internal causal role, which helps account for T's cognitive significance (i.e., the fact that a fully rational thinker may accept, e. g., the thought *Hesperus is a planet* and reject *Phosphorus is a planet*, and
- Another, "long arm" role which explains why and how T has the truth-conditions it has (and, thus, explains the identity of the two thoughts above in this respect.)

Fodor and Lepore (1992, chapter 6) point out that it is an open question how this two components are coordinated: "Why can't you have a sentence that has an inferential role appropriate to the thought that water is wet, but is true iff 4 is a prime?" Fodor and Lepore (1992, p. 171)

Now, the question whether a thought could have two different meanings according to its causal role and its truth conditions is, from the perspective of my proposal, meaningless. There is no one-one correspondence between meanings and actual causal roles, and no way to analyse one in terms of the other. I have briefly explained in 4.6.4 how facts about cognitive significance and facts about reference are explained by facts about the individuation of concepts -which depends only on the act of creation that resulted in its existence.

4.8.3 *Holism.*

The causal roles of the concept CAT will be different for any two thinkers. So, if meaning is to be equated with causal roles, no two thinkers will mean the same thing with their concept of CAT. The usual way to develop this criticism is by then pointing out that the CRS-theorist needs to distinguish a part of the causal role that is constitutive of something's having the meaning that it has, and a superfluous part that may vary among thinkers. This (again, cf. Fodor and Lepore (1992)) looks like an endorsement of the analytic-synthetic distinction, which is probably to be rejected. So, it's either holism or embracing the a/s distinction, and both alternatives are unwelcome.

An attractive answer to the objection is pointing out that *no* part of a causal role is constitutive of meaning. Rather, the causal role of the concept of F in some subject may vary enormously, but it will still be the case that the causal roles of other mental terms created by CONC will have been sufficiently varied as to make sure that, say, kind G made it the case that the changes created by CONC in answer to cue S_G were fitness-conducive.

An example may help to clarify this. Suppose that CONC works with cues that are a combination of retinal shadows and a sound -say animal alarm calls. In the presence of a combination, S_a, of these two things, CONC creates a mental term *a* that is then used to store information about an individual. If, e. g., thoughts with the content *a is around* and

F is around cofire very frequently, PRED creates a thought *a is F*, etc. The mental economy formed by CONC, PRED and the different terms they create and link has been fitness-conducive and this can be used to fix the content of future, not selected-for products of these mechanisms -see 4.4.2. There is no need that any particular subset of the thoughts in which, say, a concept of *a* participates be true of *a*. It is only needed that the concept of *a* was created by CONC -which has its own selection history, and thus helps fix the content of its products- in the presence of S_a, where this cue is related with the individual a by whatever it is that has related cues to individuals throughout selection for CONC. No analytic core is needed for this. In fact, a subject could harbour only false thoughts using the concept of *a* in question. It is still a concept of *a* if it was created by CONC in the presence of S_a.

Part II

MODALITY

5 MODAL CONTENTS

The neo-rationalist account of modal epistemology (*e. g.* Bealer (2002), Chalmers (2002), Yablo (1993)) has been extensively criticised in the last few years, but few alternative accounts of our access to modal facts have been proposed in its stead. In this chapter I lay the foundations for doing precisely that, by sketching an etiosemantic account of modal contents.

I will defend that a sizeable portion of the space of possibilities may be reconstructed as quantifications over times and probabilities. I will assume that contents about the past are easily naturalisable along the lines described in earlier chapters and will dedicate most of the chapter to show that contents about probabilities also are. For this, I discuss a simple example in which, I wish to argue, it is natural to credit a particular agent with probability-involving contents (5.2 and 5.3). As in earlier chapters, I then go on to provide a more general recipe for the attribution of ephemeral probability-involving contents (5.4). I end a first part of the chapter by putting these results in the light of the discussion in chapter 4, and sketching how concepts themselves may encode this modal information (5.5).

A second part of the chapter is dedicated to elaborate paraphrases of modal idioms in terms of probabilities and times (5.6). This may be seen as a development of Forbes's *branching conception of possible worlds*. We may, therefore, explain in fully naturalistic terms how the kind of modal contents that are constituted by probabilities and times may be thought. This completes the naturalisation of modal contents.

The rest of the chapter is dedicated to spell out some consequences of the view (5.7), including an attempt at explication of our intuitions regarding the necessity of origin, and a first stab at the naturalisation of epistemic possibility (5.8). Finally, in 5.9 I briefly take up a Humean objection to my approach.

5.1 PERCEPTUAL CONCEPTS, INDIVIDUAL AND KIND

In a recent article, Papineau (2007) has developed an account of perceptual concepts, in order to illuminate several features of (the obscurer) phenomenal concepts. I will not discuss his account of the latter, but concentrate instead in an interesting feature of his proposal regarding the former. My aim in this section will be to show that this account of how the content of perceptual concepts gets fixed cannot be completely right. The reasons why it is not will lead us towards a first stab at the naturalistic account of modal contents it is my aim to develop.

According to Papineau, perceptual concepts are the kinds of concepts we use when we make mental reference to things we have perceived. We form a perceptual concept upon our first perceptual encounter with an entity, and it is reactivated when we perceive that entity again; we also use our perceptual concepts when we imagine the entities they are about -*cf.* Papineau (2007, p. 113).

For Papineau, perceptual concepts involve a stored *sensory template*. Incoming perceptual input may, or may not, "resonate" (2007, p. 115)

with the stored template. Such reactivations, it is to be supposed, amount to (possibly mis-) recognitions of the referent of the concept. The perceptual concept is used for gathering information about its referent; when the perceptual concept reactivates, the additional information is also reactivated.

Perceptual concepts can be used to think about tokens (*There goes that dog again!*) or types (*Hey, another one of those dogs!*). What determines whether we are dealing with a perceptual kind concept or an individual kind concept? Papineau's proposal is that it depends on which kind of information is to be carried over from one reactivation of the concept to the next. For instance, if we are disposed to attach pieces of information such as "has an injured eye" to the referent of our concept whenever it reactivates, it is the concept of an individual, say, bird; if not, but we are only willing to attach information such as "has bright-coloured feathers" or "flaps wings very quickly", then it is a kind concept. The sensory templates of perceptual concepts, Papineau suggests, come with "slots" ready to be filled. Individual-persons concepts come with a eye-colour slot; dog kind concepts come with a normal-size slot, etc. Papineau (2007, p. 117)

What happens when we have both a perceptual individual concept (for some particular bird, say) and a kind concept (for the type of bird whence the former belongs)? Papineau suggests that we should regard perceptual concepts as forming *structured hierarchies*, with the individual concept adding detail to the kind concept.

5.1.1 *Modal Information.*

But things must be much more complicated, at least in the following respect. Consider again the case of our perceptual concept of a bird. We now see that it has a broken wing. The doctrine I have been just reviewing has it that, if we have both the concepts of an individual bird and of the species whence it belongs, we will attach the information that *it* is broken winged only to the individual concept -the kind concept allowing for individuals with healthy or injured wings. So far so good, but now, what should we do with the kind concept? Should we leave it as it is? Well, that would be a missed opportunity, because we *have* learned something about the species; namely, that its individual members *can* have their wings broken. In general, whenever we gather evidence to the effect that an individual has the property P, we have also gathered evidence to the effect that the members of its species *may* have that same property. This information can be very useful sometimes. If I have encountered an aggresive dog once, it is surely good to know from then on that dogs *can* be aggresive.

The problem comes now, with the part of the doctrine that talks about structured hierarchies of increasing detail, individuals being at the bottom and species higher-up. Because, many times, once we learn that an individual has some property P we can also rule out that it can have other incompatible properties. Dogs can be black but a white dog can't. That is, perceptual concepts -and concepts in general- cannot be structured hierarchies as Papineau envisages them. When I see a white dog, I am not seeing a white dog that can be black -importing information from the kind on to the individual. Some modal information -which, put in the terms of Papineau's view, is surely an important part of the information that gets attached to sensory

templates- is not of the kind that can be imported from species into individuals.

5.2 SENSITIVITY TO MODAL INFORMATION

In etiosemantics, the difference between individual and kind concepts is not a matter of the information that we are disposed to carry over, but of what type of entity explains the presence of the concept in question. If a kind, then it is a kind concept. If an individual, then it is an individual concept[1].

This simple account, it seems, runs into troubles when facing the issue of building sensitivity to modal properties; at least when such sensitivity is to be understood in the broadly causal-informational way I am favouring. It is commonly assumed throughout the literature that such a sensitivity would be entirely mysterious. So, for instance, Peacocke suggests that it would need the postulation of "dubiously intelligible faculties connecting the thinker with some modal realm" Peacocke (1999, p. 163). Another example of this stance, from the other end of the philosophical spectrum, is Millikan denying that true negative sentences have as their meaning nonexistent world affairs, because "nonexistent world affairs would surely have no powers in the causal order, hence could not play roles in Normal explanations" Millikan (1984, p. 221). The most important task of this chapter will be to show that sensitivity to modal properties is perfectly intelligible, although, of course, it involves no appeal to the causal powers of the uninstantiated.

5.2.1 *Frogs, Goodflies and Badflies*

Let me introduce a complication in the example of Democritus the frog and its fly-involving mental states. Now, not all of the flies that live near the Pond where Democritus and his conspecifics hunt are nutritious for them. In fact, only a 30% of them are -we can call these *goodflies*. The rest are harmless, but they provide no advantage to the frogs who eat them -and we will call these *badflies*.

As in previous chapters, we can study the situation for the denizens of the Pond from a cost-benefit perspective: hunting for flies has a cost in fitness -because of resource expenditure-, and only securing the capture of goodflies yields a benefit in fitness. The fittest individuals will be those that, in the long run, reproduce differentially better, and will pass their strategies onto ulterior generations.

To simplify the discussion we can assign a numerical value to the costs and benefits in resources in which hunting frogs incur, and assign probabilities of success to the different possible courses of actions frogs can take, in the different circumstances. I will consider three different strategies followed by three different frogs, Empedocles, Epicurus and Democritus. I will furthermore help myself to the implausible but simplifying assumption that these hunting strategies have simply popped

1 There is a further, syntactical difference among these types of concepts. For example the input to the predication producer PRED (see chapter 4) was supposed to be an individual concept and a kind concept, and this presupposed a means to distinguishing both syntactically. We may suppose that further complications in the conceptual setup of an individual will bring these types of concepts apart as regards their syntatically acceptable positions in a thought. Following Papineau, I am leaving these complications aside here.

into existence because of mutation in the frogs, instead of through a process of selection:

- Democritus has a mechanism N that goes *on* whenever he sees a fly -regardless of whether it is a goodfly or not[2]. N, in its turn, causes Democritus to protract his tongue. The net result is that, whenever a fly is near Democritus, he hunts. His success rate in hunting is 0,8. That is, of every 10 attempts at hunting, he successfully secures the prey an average of 8. Democritus has such a high success rate because he doesn't leave the fly time to react. As soon as he sees it, he protracts his tongue[3].

- Epicurus has a mechanism M that only reacts to goodflies. When a fly first comes flying by, he does not even notice it, but once he sees that it is a goodfly (M's input is a tiny red dot that only goodflies have in their abdomen[4]) he hunts -M causing the protraction of the tongue. The hunting success-rate for Epicurus is 0,6. It is somewhat lower than Democritus's because it takes some time to see that something is a goodfly (the red dot is really tiny, and not easy to spot) and so, by the time that Epicurus has found out, the goodfly is normally starting to move away from the frog. Cost and benefits of hunting are as with Democritus.

- Empedocles has a slightly more complicated strategy. When he sees a fly a mechanism we could call PREPARATION goes *on*. PREPARATION's being *on*, in its turn, kicks off a number of changes, *e. g.*, increased hearbeat rate, increased attention and the like, that serve as a preparation for hunting: if a frog hunts when PREPARATION is *on*, its probabilities of success are increased.

  Besides, Empedocles has a mechanism M that is an exact copy of Epicurus's: when he sees the red dot that is the tell-tale sign of goodflies, M goes *on*, and this causes Empedocles to hunt. The difference is his success rate -PREPARATION being *on*- which is as good as Democritus's: 0,8. On the malus side, the cost of PREPARATION's going *on* is -5 resource units (*rus*, from now on).

I will also assume that, for the three frogs, the cost of hunting (whether successful or not) is -20 rus; the benefit of catching a goodfly is 500 rus; the benefit of catching a badfly is 0 rus -*cf.* figure 11. We may now calculate the Fitness Contribution of their hunting systems. We will disregard the values outside the diagonal of the Fitness Matrix, given that our simplifying assumptions (*i. e.*, that they are perfect detectors) imply that the frogs will never cash them:

2 For the sake of simplicity I will be assuming that Democritus, and the other frogs, make no mistakes in recognising flies. That is, that the probability of there being a fly given that the input to Democritus's M is present is 1.

3 To put this in the context of the discussion in previous chapters, we are assuming here that Democritus is somehow perfect in his detection of flies. This is, strictly speaking, impossible: Democritus will always be using a proximal cue which will never have a perfect, counterfactual-supporting correlation with flies. First, there is the empirical reason that no proximal cue will, as a matter of fact, be such a perfect tracker. Then, there is the theoretical reason that -as I have argued in 2.5- the property of *Being such-and-such a natural kind* is not a Shoemakerian property, which is what detectors detect.
So, the *success rate* in the main text does not refer to Democritus's ability to identify flies -that is, this rate is none of the probabilities in the $\mathrm{IP}_{\mathtt{fly}}$, which we are assuming identical to the identity matrix. The success rate, instead, is useful in calculating $\mathrm{FM}_{\mathtt{fly}}$: how big a prize is to identify a flie, given Democritus's subsequent ability in hunting for it.

4 Here again, for the sake of simplicity, I will assume that Epicurus's $\mathrm{IP}_{\mathtt{goodfly}}$ is the identity matrix.

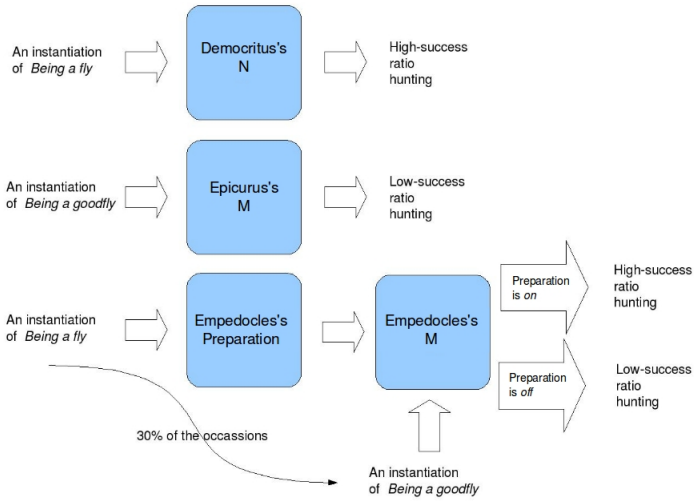

Figure 11: Three Strategies

- For Democritus, the fitness value associated with a correct positive is $w_{11}^{fly} = -20 + 0,8 \cdot 0,3 \cdot 500 = 100$, and a correct negative $w_{22}^{fly} = 0$.
- For Epicurus, the values are $w_{11}^{goodfly} = -20 + 0,6 \cdot 500 = 280$, and $w_{22}^{goodfly} = 0$.
- Finally, for Empedocles, $w_{11}^{fly} = -5 + 0,3 \cdot (-20 + 0,8 \cdot 500) = 109$ and $w_{22}^{fly} = 0$.

And the overall Fitness Contributions for each of them:

- $FC^{Democritus} = 100 \cdot P(fly)$
- $$FC^{Epicurus} = P(fly)(P(goodfly|fly)/P(fly) \cdot 280) = P(fly)(0,3 \cdot 280) = 84 \cdot P(fly)$$
- $FC^{Empedocles} = 109 \cdot P(fly)$

Let us see the three frogs in action in a little simulation. The game goes as follows: in each round (*i. e.*, hunting episode) a fly gets near each of the frogs. We are assuming that there is no direct competence between them. That is, that a frog catches the fly is just a matter of the effectiveness of its strategy. That the fly is good or bad is simply a matter of chance (with the probability of flies being goodflies being 0,3) but, in each round, all three flies facing each of the frogs will be of the same kind. When faced with a fly, Empedocles, Democritus and Epicurus will do as described above, and the amount of resources they cash or loose will depend partly on their behaviour (PREPARATION's going *on* or not, for example), and partly on the success rate of their hunting.

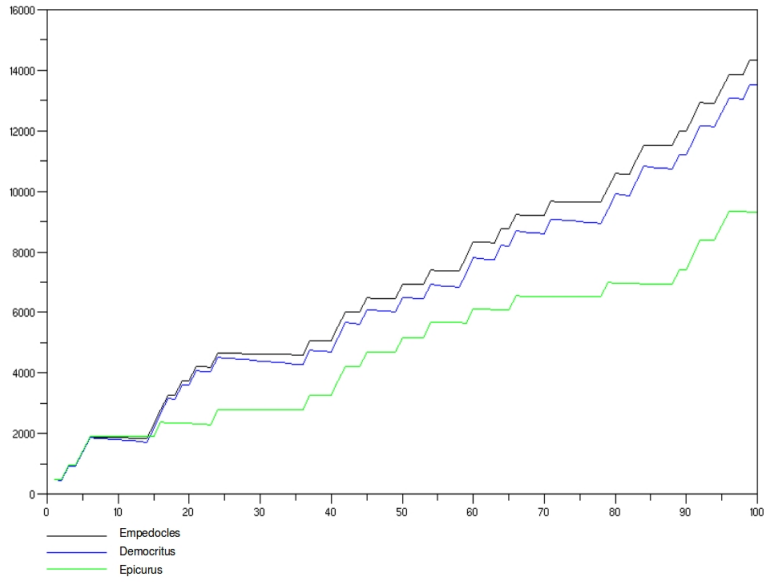

Figure 12: The three frogs after 100 hunting episodes

After 100 hunting episodes, the situation is as shown in figure 12. In the x-axis we have the number of hunts, and in the y-axis the net resource-balance for Empedocles (black line), Democritus (blue) and Epicurus (green). First, we can see that Empedocles's and Democritus's resources grow parallel, although Empedocles's are increasingly higher. This is because they have success in exactly the same hunting occasions -their success rate is the same- but, where Democritus always spends 20 rus -he hunts each and every fly-, Empedocles does better and only spends the 5 rus of PREPARATION's going *on* in the cases in which he will end up not hunting -because the prey is not a goodfly. As for Epicurus, he doesn't even spend the 5 rus of PREPARATION so, if he were to be as successful in hunting as Empedocles or Democritus, he would outperform them. But he is not; his lower success-rate in hunting makes all the difference -the green line makes less "jumps up" that black or blue. After ten thousand hunts (Figure 13) the differences only grow bigger.

This is simply a more vivid presentation of the information already present in the Fitness Contributions: Empedocles's strategy is the most profitable in the long run. He would be fittest and would be the one to reproduce and pass his strategy on; he would be selected for. The situation with Empedocles is not much different from the situation with Democritus in DEMOCRITUS AND THE CONTENT OF M'S BEING ON, in 1.3: Empedocles has a couple of mechanisms, M and PREPARATION, that indicate a number of properties, and this has made him more successful than conspecifics with a different cognitive setup. This is the kind of story that earns content attributions for its main characters. Let us, then, work out the correct attribution of content to these mental mechanisms's being *on*. We will see that there *is* a state with a probability-involving content, but that it is neither M nor PREPARATION's being *on*. It is, instead, the state consisting in these two mechanisms's being wired the way they are.

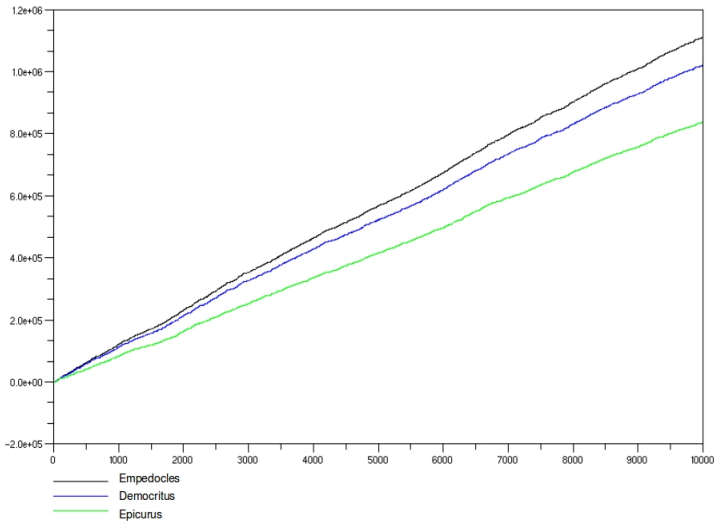

Figure 13: After 10000 hunting episodes

5.3 CONTENT ATTRIBUTION TO EMPEDOCLES'S MENTAL STATES

5.3.1 *M and Preparation*

What is the content of the positives of M and PREPARATION? We may use the THERE IS AN F AROUND recipe for content attribution introduced in chapter 1: I argued there that the content of the positives of an atomic indicator mechanism should involve the HPC that explains the correlation between the Indication Profiles and Fitness Matrices of that mechanism. If so, M is easy: its success relies, in the particular way we reviewed in chapter 1, on the existence of a Homeostatic Property Cluster that explains the frequent coinstantiation of properties that are close to its Input (maybe *Being a red dot in the abdomen of a fly*) and other properties (such as *Being nutritious for frogs*) that explain the fitness-conduciveness of reacting to properties in the input. The HPC in question, we are assuming, is *goodfly*. So the content of M's being *on* is *There is a goodfly around* univocally -or so I argued in the analogous case of chapter 1.

What about PREPARATION? Its input are properties in the fly Property Cluster, but the explanation of the values in the Fitness Matrix is slightly more complicated than other cases we have seen so far. We calculated above the overall fitness of a correct positive for Empedocles: $FC^{Empedocles} = [-5 + 0,3 \cdot (-20 + 0,8 \cdot 500)]\, P(fly) = 109 \cdot P(fly)$. This fitness is a joint merit of the workings of M and PREPARATION. M gets the most benefitial fitness contribution:

$$FC^{M} = [0,3 \cdot (-20 + 0,8 \cdot 500)]\, P(fly) = 114 \cdot P(fly).$$

Apparently, the M in Empedocles does a much better job than the M in, for example, Epicurus. But this comes down to its better success rate when hunting, and we know that the rise in success rate is facilitated by PREPARATION. In creatures such as Epicurus, which lack a token of

PREPARATION, success rate is significantly lower. PREPARATION makes Empedocles enter in a state that, *if the fly is a goodfly*, will help hunt it more efficiently. This is Empedocles's competitive edge against Epicurus, who lacks this preparation. Besides, the cost of causing this state is low enough so that, *if the fly is not a goodfly*, losses are minimised. This is Empedocles's competitive edge against Democritus, who incurs in the full cost of hunting every time. PREPARATION steers a middle course between fully responding to flies as Democritus, and ignoring them as Epicurus. PREPARATION makes Empedocles a *more cost-effective* hunter. It is easy to see this if we compare his resource-curve with Democritus's. The only difference between both curves is that Democritus's goes down 20 rus everytime a fly approaches him, while Empedocles manages to spend only 5 rus in the event of an encounter with a badfly.

PREPARATION is exploiting a homeostatic mechanism that makes fly-cluster properties come together, the very mechanism that explains that a 30% of the times the cluster includes the nutritional properties of goodflies[5]. This homeostatic mechanism is the one that individuates flyhood. So, PREPARATION's positives have the content *There is a fly around*, just as M's positives have the content *There is a goodfly around*.

None of these two contents reflect the fact that $P(\texttt{goodfly}|\texttt{fly})$ is crucial to the success of Empedocles's strategy versus Epicurus's and Democritus's. It appears, though, that such a probability should be part of a content-endowing explanation. The question is, the content of *which* state?

5.3.2 *Synergic Associations*

The relative "positions" of M and PREPARATION in Empedocles's cognitive setup bear some resemblance to the relative "positions" of F- and G-mechanisms in the cases of conjunctive and disjunctive recruiting I discussed a couple of chapters ago, in 3.1. There, too, the Fitness Contributions of both mechanisms were mutually dependent -one could see the process by which the F- and G-mechanisms came to be associated either as the F-mechanism recruiting the G-mechanism for its output, or as the G-mechanism recruiting the F-mechanism for its input, and the gain in Fitness Contribution harvested by their association could be allocated to the account of either of the two mechanisms.

I wish to suggest that the state formed by PREPARATION, M and their association[6] has a probability-involving content. I will call this kind of cost-effective associations among mechanisms, *synergic associations*. A probability-involving content attribution to synergic associations is sensible: for example, it is the probabilistic relation between instantiations of the property of *Being a goodfly* and those of the property of *Being a fly* that explains the success of the synergic association of M and PREPARATION. It is in the spirit of my former discussion of conjunctive and disjunctive recruitments to claim now that the association of the two mechanisms has the content that there is whatever causally-grounded relation between Fs and Gs that explains that, if there is an F around, the probability of there being a G around is in a certain interval:

5 By the way, there need not be any fundamentally random mechanism at bottom of such probabilities; for all Empedocles and Democritus care, the homeostatic mechanism may produce good- and badflies in the perfectly deterministic sequence {b,g,b,b,g,b,b,g,b,b}.

6 That is, the causal underpinnings, whatever they are, of the fact that PREPARATION's being *on* improves the Fitness Matrix associated with M, at a low cost.

PROBABILISTIC RELATION: An agent A has a state, N, with the content *There is a causally-grounded relation between Fs and Gs such that* $x > P(G|F) > x'$, if the following explains that N exists

1. A has a mechanism whose positives have the content *There is an F around* and another with the content *There is a G around*.
2. N is a causal association of the F- and the G-mechanisms alluded to in 1 such that $FC_F + FC_G$ when the states are associated is higher than $FC_F + FC_G$ when they are not.
3. The difference in fitness contributions in 2 is explained by whatever causal underpinnings the fact that $x > P(G|F) > x'$ has.

We may have qualms about attributing a probabilistic-relation content on such a meager basis. One of the reasons for this is that it is impossible to read the probability the mental state has as content off its causal profile: there is no one-one relation between, on the one hand, the kind of relation between F and G (probabilistic or otherwise) and, on the other hand, the form of the synergic relation between F-mechanism and G-mechanism. A consequence of this is that many of the ways of associating two mental mechanisms we have been studying -*e. g.*, disjunctive/conjunctive recruitment in Chapter 2 and the kind of synergic relation that holds between PREPARATION and M- may be useful in overlapping cases. If, *e. g.*, the proportion of flies that are goodflies is high enough, simply effecting a disjunctive recruitment of PREPARATION for M's input may be enough to make the difference with other competing agents that have not stumbled upon the recruiting. Even more: if the synergic relation is not very benefitial -*e. g.*, if the cost of PREPARATION's going *on* is too high, or the increment in the hunting efficacy of M too low- the disjunctive recruiting may yield a *higher* increase in FC than the synergic association.

We encountered the same situation in chapter 3, when discussing states that represent causal relations: there is no straightforward correlation between the causal profile of the state and whatever it is that it represents. This correlation appears only when there is a second order mechanism (LTP* was our example there; see 3.4) that produces states that represent causal relations.

In the following section, again, we will see that, when there is such a producer mechanism that outputs states that represent a probabilistic relation among properties, we find some correlation between form of the state and type of relation that it represents. This is an important feature, because it is the causal profile of particular states, and not the historical properties upon which content supervenes, that other states see. That is: once we move pass the simplest train of thoughts, we need some level of isomorphism between the probability a state represents and the causal profile of this state. This is what we will attain in the following section.

Before that, though, I wish to discuss an interesting special case in which the probability in question may be read off the causal dispositions of the contentful mental state -even if the latter is selected-for.

Evolutionarily Stable Strategies.

A situation such that a disjunctive recruitment is chosen over and above a synergic association in the face of two probabilistically related

properties will not happen if we allow all possible mental mechanisms -under certain restrictions- to appear through mutation and fight to prevail.

For a concrete example, in the three-frogs scenario we have been discussing, we have seen that, at zero cost, you only get a 0,6 success ratio, which goes up to 0,8 if you spend 5 rus. We may assume that these two values are connected by a function that raises the success ratio with every raise in the cost of PREPARATION. A reasonable cost function would approach success-ratio of 1 asymptotically as cost progresses towards infinite. One such function[7] (that, besides, yields 0,6 success-ratio when cost is 0 and 0,8 success-ratio when cost is 5, to keep the example continuous with the previous discussion) is, for example,

$$\text{Success Ratio} = 0,6 + \left(0,4 - e^{\frac{\ln(0,2)}{5}\text{cost}}\right)$$

Thus, the fitness contribution of the different strategies, taking this cost-function into account would be:

$$\begin{aligned} \text{FC}(\text{cost}) \quad = \quad & \Big[-\text{cost} + 0,3 \cdot \\ & \left(-20 + 500 \cdot \left(0,6 + \left(0,4 - e^{\frac{\ln(0,2)}{5}\text{cost}}\right)\right)\right)\Big] \text{P}(\text{fly}) \end{aligned}$$

This formula is such that $\text{FC}(0) = \text{FC}^{\text{Epic.}}$ and $\text{FC}(5) = \text{FC}^{\text{Emped.}}$[8]. $\text{FC}(\text{cost})$ has a maximum around $\text{cost} = 11,3$. The strategy that implements a preparation with that cost is *evolutionarily stable* in Maynard Smith's sense (see Maynard-Smith (1999)): no different strategy (that still complies with the cost-constraints stated, that is) can penetrate a population implementing such an EES -see figure 14.

In the ideal situation in which, given a particular fitness landscape such as that defined by the three-frogs example, there is enough time and variation for natural selection to reach the EES, we could, so to say, read $\text{P}(\text{goodfly}|\text{fly})$ off the preparation cost in the strategy. This situation is, maybe, most confortable for the content internalist's intuitions, in that it allows for a reconstruction of the external property in question (the conditional probability, in this case) in terms of features of the internal system -the cost/success-ratio pair chosen by the winning strategy. But it is important to remember that such a situation may be unrealistically idealised: evolution may not have time to find the EES, or it may be inaccessible for reasons having to do with the gene pool -*cf.* Bell (2008, chapter 3). Even if the final strategy is not in the maximum of the fitness landscape, it is still the case that one particular conditional probability -the one that happens to be the causally-supported one in the context- figures in the explanation of the survival of the suboptimal, but good enough, winning strategy. This is why the conditions for a selected-for mental state to have the content *There is a causally-grounded relation between Fs and Gs such that* $x > \text{P}(\text{G}|\text{F}) > x'$ are not very informative.

It would be wrong to take these considerations as pointing to another constraint for our content attributions:

7 Among many possible alternatives. This is just a whimsical example.

8 To include Democritus in the picture we would need to complicate this formula slightly. The point I am after can be made making reference to Epicurus's and Empedocles's strategies alone, though, so I will not tax the reader's attention with further niceties.

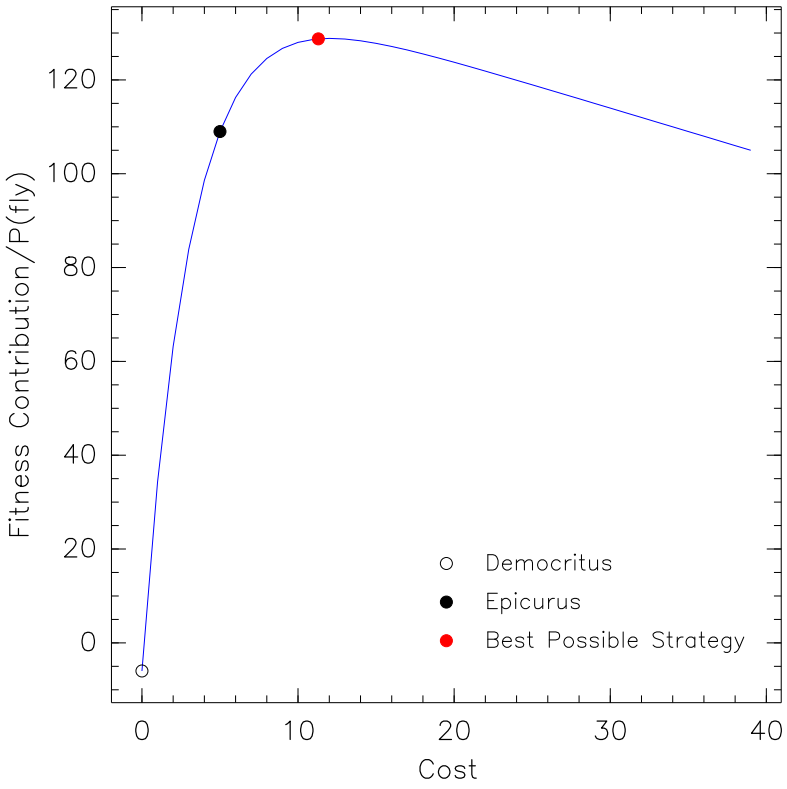

Figure 14: The fitness landscape

INFORMATIVENESS: If some state S in an agent A has a content C, it must be possible to reconstruct a satisfactory amount of C out of the role of S in A.

Informativeness is an internalist requirement. An extreme version of this constraint is,

CAUSAL ROLE: A state S in an agent A's having a content C consists of S's having a concrete role R in A.

I trust that the reader will have found, in the discussion in chapter 4, enough reasons to reject both INFORMATIVENESS and CAUSAL ROLE.

Informativeness expresses the feeling that, even if we are not ready to equate content with position in a conceptual network, there must be some interesting portion of the content of a mental state that may be recovered by observing the actual workings of S -regardless of its causal-historical properties. This is related to, but different from, another internalist requirement that could be made using the terminology introduced in chapter 3,

FEW BLAMELESS WRONGDOINGS: The Procedure a producer mechanism follows must provide a reliable guide to the content of its products; the amount of blameless wrongdoings must be adequately low -and this *adequately* is not merely *adequately**, that is, not merely to be explicated in terms of Fitness Contribution balances, but follows some independent norm.

We should reject both INFORMATIVENESS and FEW BLAMELESS WRONGDOINGS. There is no independent norm to help us gauge the

informativeness or the blameless wrongdoings a selected-for mental state is allowed to make; this constraint introduces an extraneous, non-naturalistic normative element. On the other hand, as we are about to see -and as we have already seen in a number of parallel contexts- if the synergic association in question is produced by a selected-for mechanism which has a function to yield such associations, an independent, fully naturalistic norm emerges that meets the intuitive informativeness constraint. I turn to show this now.

5.4 PRODUCERS OF SYNERGIC ASSOCIATIONS

I shall start from a just-so story about a possible mechanism which has been selected for producing synergic associations; after that, I will give more general conditions for the attribution of content in this situation.

Consider an agent A such that there is a magnitude mg (*e. g.*, A's hearbeat-rate, or degree of attention) with the following characteristics: First, it has a base value, mg_o. Second, it is possible to increase or decrease mg, with the corresponding expenditure or saving in resources; that is, the value of the increment of mg is a monotonically-increasing function of its cost, $mg - mg_o = \Delta mg = f(cost)$. Finally, the increase on fitness values, w_{ij} of some mental mechanism M of A's, whose positives have the content *There is a G around*, is a monotonically-increasing function of Δmg, thus: $\Delta w_{ij} = g_{ij}(\Delta mg) = g_{ij}(f(cost)) = w'_{ij}(cost)$.

So, the gain or loss of total Fitness Contribution of M, after devoting a certain amount of resources, *cost*, in raising *mg* is:

$$\begin{aligned}\Delta FC_G(cost) = {} & -cost + \\ & P(G)\left(w'_{11}(cost)\,P(on|G) + w'_{12}(cost)\,P(off|G)\right) + \\ & P(\neg G)\left(w'_{21}(cost)\,P(on|\neg G) + w'_{22}(cost)\,P(off|\neg G)\right)\end{aligned}$$

We could envisage a mechanism N that works according to the following Procedure[9]:

PROCEDURE - SYNERGIC: Whenever two mental mechanisms of A, an F-mechanism and a G-mechanism, are such that

$$P(G - state\ is\ on|F - state\ is\ on) = X$$

if $cost_{max}$ is the value that makes $\Delta FC_G(cost)$ maximum, then add the following action to the output of the F-mechanism: Spend $cost_{max}$ in increasing mg.

N, following PROCEDURE - SYNERGIC, effects a change in the F-mechanism that links it with the G-mechanism: now the F-mechanism affects the performance of the G-mechanism in a way in which it previously did not, and the result is a higher overall Fitness Contribution. N would be more nuanced in its behaviour that, for example, LTP* in 3.4, in that it would not simply create all-or-nothing recruitings among states, but rather taylor the association of states it produces to reflect the probabilistic relation that its target properties have in ther real world. Given the very tangible contribution that N makes to the fitness of its possessor (by improving the Fitness Contribution of the states it associates) it is possible that it had a non-negligible influence in the

9 N needs to use a proximal cue to know about the conditional probability below. I'm leaving this complication aside.

survival of its possessor and that, by that token, became fixated in its lineage.

The presence of a higher-order producer mechanism such as N helps alleviate the uninformativeness of conditions such as PROBABILISTIC RELATION above: in the case of mechanisms that comply with PROBABILISTIC RELATION and are not the products of a previously selected mechanism, one cannot recover the information about the probabilistic relation among instantiation of properties around the possessor of the mechanism out of the workings of that very mechanism -barring assumptions to the effect that the environment would only allow an Evolutionarily Stable Strategy to get fixated. That is, pretty much *any* position in the fitness landscape in figure 14 could be selected because of its sensitivity to probabilistic relations among properties, if no other mutant has been lucky enough to stumble upon a better strategy -one nearer the maximum, that is.

This is not the case with states created by selected-for producers such as N, not even for mechanisms that are less accurate in calculating the most cost-effective investment in magnitude mg. Any such producer will produce several states, that represent different probabilistic relations among properties. From such a set of states, and if we have some, even partial, information about the probabilities they are representing, we may be able to "reverse engineer" the procedure the producer uses to create them. This would provide a route from the state and its characteristics back to the probabilities it involves in its content. This, in turn, will help meet INFORMATIVENESS. More importantly, other states in agent A may make themselves sensitive to the systematic relation between causal profiles of mechanisms created by N and conditional probabilities in the environment.

5.4.1 *A more general set of necessary conditions.*

Now, in general, what are the conditions for some mental mechanism N to be such that the associations it effects have a probability-involving content? The following set of conditions (closely related to the analogous set of conditions in 3.8) do the trick, for a mechanism that takes an F-mechanism (*i. e.*, a mechanism whose positives have the content *There is an F around*) and a G-mechanism and effects a synergic association with the content *There is whatever causally-grounded relation between Fs and Gs that explains that* $x > P(G|F) > x'$.

> Inputs of type C, such as C_m, cause N to create synergic associations among mechanisms $f_a(C_m)$ and $f_b(C_m)$. The form of the association is also a function of the cue -say, $f_{assoc}(C_m)$. We'll call this association $SA_m(a, b)$.

This is a substitutions of variables in the ROUTINE schema we saw in 3.8. Besides,

> A sufficient* number of the synergic associations SA_i produced by N in the presence of cue C_i have been fitness-conducive because
>
> - Cues such as C_i are a good enough sign of the fact that the HPCs *a* and *b* -which are the ones that the content of $f_a(C_i)$ and $f_b(C_i)$ involve- are in a probabilistic relation such that $x_i > P(a|b) > x_i'$ -the interval of probabilities is also function of the cue.

- The added Fitness Contributions of the a- and b-mechanisms after their synergic association in $SA_i(a,b)$ has been higher than before their association, and
- The fact that $x_i > P(a|b) > x_i'$ figures in a relevant explanation of why 1 is the case[10].

This is analogous to FITNESS CONDUCIVENESS in 3.8. The idea is that what explains the fitness conduciveness in each succesful product of N is a mechanism similar to the ones appealed to in PROBABILISTIC RELATION above. Finally, as always, we need to make sure that the workings of N are covered by an HPC that keeps together cues and synergic associations:

> There is a higher order HPC Q^n that explains that instantiations of properties C_i go together with the fact that $x_i > P(a|b) > x_i'$ where *a* and *b* are the HPCs targetted by mechanisms $f_a(C_i)$ and $f_b(C_i)$ respectively.

This is analogous to HIGHER ORDER HPC in 3.8. Once we have a mechanism N with such a causal profile, embedded in such a causal net, we can use it to create states with probability involving contents:

> An agent A has a state, $SA_i(a,b)$, whose positives have the content *There is a causally-grounded relation between as and bs such that* $x_i > P(a|b) > x_i'$ if cue C_i has caused N to create it.

The basic idea is the same as in previous chapters. The mechanism that is actually selected for, N, has stumbled upon a causal structure that effects a correlation between a property of type C and the fact that two natural kinds F and G are such that $x > P(G|F) > x'$. This has been useful because N's outputs are states that take profit of just that kind of probabilistic relation among natural kinds -states such as the one that Empedocles uses to beat Democritus and Epicurus in the fly-hunting competition.

A perfectly intelligible faculty.

A mechanism that follows, say, PROCEDURE - SYNERGIC is going to be pretty sophisticated, to be sure. For starters, it has to be able to find the maximum of $\Delta FC_G(cost)$, which in turn involves solving the equation $\frac{d\Delta FC_G(cost)}{dcost} = 0$, and that would be a pretty impressive feat for a mutation to achieve in one step. Naturally-evolved mechanisms that implement something similar to this procedure will derive from other, less accurate mechanisms -ones that give a coarser approximation to $cost_{max}$, for example. In any event, even if it is implausibly sophisticated, there is one thing it is not: it is not the embodiment of a "dubiously intelligible faculty", as in Peacocke's turn of phrase. It is perfectly intelligible how and why could an organism go about following PROCEDURE - SYNERGIC. And yet, as I will try to show in the following sections, this is the basic building block for our modal competence.

10 A "relevant" explanation will be relevantly similar to the explanation of Empedocles's success in the three-frogs example.

The relation between F-Mechanism and G-Mechanism.

It should be noticed that the synergic association between the F-mechanism and the G-mechanism need not be a direct causal relation. It is compatible with PROCEDURE - SYNERGIC that the effect of the investment in *d*mg occurs *downstream* from the G-mechanism. What N does is making the F-mechanism cause some changes that have nothing to do with the workings of the G-mechanism, which it leaves untouched.

In such a case, we should describe the situation as one in which the F- and G-mechanisms jointly cause a (number of) event(s) with a positive Fitness Contribution for A. Of course, in other cases PROCEDURE - SYNERGIC may create a *bona fide* causal relation between the two mechanisms, wherein, for example, the G-mechanism changes as a result of the F-mechanism's going *on*. This very simple setup may help see a particular family of problems under a clearer light. What makes the F-mechanism and the G-mechanism produce representations that are somehow linked may not be a direct causal connection. But this does not mean that mere co-activation of the two mechanism is enough for their representations to be so-linked. Such a theoretical possibility may be an interesting avenue to a solution to the binding problem in psychology -*cf*. Revonsuo (2009) and references therein.

I should also quickly point out that implementing a *probabilistic* causal relation between the F-mechanism and the G-mechanism in response to a probabilistic relation between Fs and Gs would be a pretty bad strategy. It is not that the activation of the F-mechanism should cause the activation of the G-mechanism a 30% of the times, if the probabilities that govern the causing are independent from those that govern the relations between flies and goodflies. Such a probabilistic causal relation would get a correct positive of the G-mechanism in only a $30\% \times 30\% = 9\%$ of the cases in which it is activated.

It is likely that a real life LTP* works more or less this way: whenever there is co-firing of F-mechanism and G-mechanism, LTP* makes the probabilistic causal relation between both a bit higher until, after sufficient co-firing, it reaches the strongest possible causal link between states. With independent firing, instead, a complementary mechanism LTP* (standing for Long Term Depression) diminishes the causal link.

Anyway, if we make an LTP*/LTD* pair work on a probabilistic relation between properties (they make, *e. g.*, the causal link strength move up and down a sigmoid curve between 0 and 1, with 0 being no causal relation at all, and 1 causal necessitation), it can be shown that, for probabilities above 0,5, the pair makes the connection reach 1 in the long run, and for those below 0,5 it makes it go down to 0. So, for high correlations you get Democritus, and for low correlations you get Epicurus. There is a better strategy, as we already know, and one that can't be mimicked just with LTP*/LTD*: Empedocles's.

5.5 CONCEPTS AND PROBABILITY-INVOLVING CONTENTS

We may now come back to Papineau's perceptual concepts. Let us recall from chapter 4 the mechanism CONC, which is able to produce concepts. CONC takes a cue S_a as input and issues a concept which may be used in thoughts of the form *Fa* -that is, thoughts in which a property is predicated of its referent.

I have criticised Papineau's suggestion that the difference between individual and kind concepts is the kinds of information one is willing to carry over from one encounter with the concept's referent to the next. As an extra reason for doubting that this is the right account, we may note that there is no limit to the amount of error that a cogniser can make, as regards the kinds of properties that may be carried over upon encounter with kinds or individuals. Someone may be (very wrongly) disposed to judge that birds have broken wings after having seen a bird with a broken wing, or that math teachers have a sunny disposition, after an encounter with a particularly nice member of the group. All of this is compatible with his having the kind concept BIRD, and the kind concept MATH TEACHER.

The right account, I urged in section 4.4.2, has it that someone has the kind concept BIRD if a mechanism such as CONC has been caused by a cue S_{bird} to produce in her a mental mechanism M, such that S_{bird} and birdhood are related by the same higher order HPC that explains the emergence of CONC -see the referred section for details. This may well be compatible with the concept in question following a comparatively lousy Procedure of application and, in particular, with the cogniser's being disposed to carry over broken-wing information from bird to bird.

But the problem with Papineau's account that most interests me in the context of this chapter is the one having to do with carrying over modal information from individuals to kinds. As I said above, the inference taking from

- Fido is ill-tempered

to

- Dogs may be ill-tempered

is perfectly safe -and may help to keep you safe on occasion. There is a straightforward way in which a mental architecture with contents fixed in the way I am advocating may implement this kind of inferences. We can envisage a *synergic associator*, SYN, which works in the following way:

SYN - CAUSAL PROFILE:
- The cue that causes SYN to work is PRED forming a thought that predicates a property *F* of an individual *a* -such as *Fido is ill-tempered*; see 4.4.2 for details on PRED.
- It then issues an order for a synergic association to be effected between a concept that is a function both of *F* and a^{11} (DOG, in our example) and the concept of F (ILL-TEMPERED, in our example).

11 To decide that it is the concept DOG, and not, e. g., the concept BROWNISH MOVING THING which must be synergic-associated with ILL-TEMPERED is a complicated cognitive task, about which much must be said -and about which I will say next to nothing. One may hypothesise with Papineau that kind concepts do carry empty slots -although, as I have been arguing, these slots provide no criterion of kind-conceptness- and that the synergic association with the concept ILL-TEMPERED is promoted in all kind concepts to which the individual belongs such that they have the right kind of empty slot. So, maybe, DOG and MAMMAL have, and BROWNISH THING has not, a *mood* slot and this rules that synergic associations with ILL-TEMPERED are promoted in the former but not the latter. But this is all wild speculation.
In any event, just how SYN solves the problem of promoting useful inferences from individuals to kinds is an enormously interesting empirical question, but one that I need not care about at the moment. For the point I'm making it is enough that the problem *can* be solved in one way or other. We are a living sign that it can.

- The strength of the association will depend on the amount of information available. In some cases, it is possible that SYN simply chooses a baseline degree of association -corresponding to a baseline degree of probability of, say, dogs being ill-tempered. In more sophisticated cognitive setups, the association effected will depend on the circumstances of the formation of the thought that Fido is ill-tempered -*e. g.*, whether its behaviour was provoked or not.

As always, to make SYN capable of endowing content to the synergic associations it effects among concepts, we need it to have obtained the function to behave as detailed in SYN - CAUSAL PROFILE in the presence of the adequate HPC. In this case, there needs to be causal grounds for the fact that cues of the type (PRED's creating a thought *Fa*) go together in a sufficient number of times with the fact that the tokens of some of the kinds[12] whence *a* belongs have a certain probability of being F.

The presence of SYN, if its function emerged in the presence of the right causal grounds, will be enough to make the concept DOG carry the (fully intentional) information that dogs are ill-tempered with some probability[13]. Let me point out a couple of the advantages of this way of accounting for the encoding of probabilities in our concepts, if you are of a naturalist persuasion:

Again, there is no mysterious appeal to uninstantiated events, or the content-endowing capabilities of the merely possible. The synergic associator simply works on causally-grounded correlations it has stumbled upon. These causal grounds are operative not only in the actual correlations SYN has encountered, but also in innumerable others. This is how, upon one sole encounter with a cue, SYN may go on to rightly promote a synergic association among concepts. Put from the point of view of the agent, upon once coming to believe that *Fa*, she then goes[14] on rightly to connect the concepts of several kinds whence *a* belongs to a certain probability of their being *F*.

Nor does the way in which SYN works involve undischarged appeals to *a priori* knowledge. SYN follows a Procedure, described in SYN - CAUSAL PROFILE and this procedure gives sufficiently* reliable access to information about probabilities. There is no realm of the *a priori* involved[15]. I will try to explain now how the synergic association of concepts effected by the likes of SYN may underlie our very own ability to modalise.

5.6 IN SITU POSSIBILITIES

We have just seen how a mechanism such as SYN may help a creature gather modal information about a kind. A similar procedure may effect the gathering of modal information about individuals. There is an apparently sound inference that goes from:

12 See footnote 11.

13 More strictly: information that there is whatever causal grounds that explain that $x > P(ill-tempered|dog) > x'$.

14 We should not think that she does this voluntarily, out of her own will, etc. These kind of mental acts have much more stringent sets of conditions of existence. All that the agent is "doing" in the main text is automatically forming a conceptual relation upon coming to believe that *Fa*.

15 Or, maybe, the realm of the *a priori* is the realm of the selected for being sufficiently* reliable. I don't find this suggestion totally unattractive.

- Fido is angry now

to

- Fido may be angry

I do not intend the conclusion to be read as following logically from the premise. That is, the argument is not supposed to rely on an instance of the axiom T which allows to infer $\Diamond p$ from p. This inference is of no practical (ecological) interest and, as such, is unlikely to be at the base of our modal competence. Knowledge of the logical law in question, and particular instances thereof, true as they are under natural assumptions, provides no competitive edge outside the Philosophy department.

Instead, what seems to be ecologically very useful is knowing that Fido may be angry again *in the future*. That is, that Fido is such that there is a non-negligible, objective probability that he be angry again in the future. In this way, we may adjust our behaviour (in the nuanced, synergic way I we have sketched) to a certain probability of certain future time slices of Fido's which we may encounter being angry[16]. We may easily imagine a mechanism, SYNIND, that works just like SYN, but encondes modal information about individuals:

SYNIND - CAUSAL PROFILE:
- The cue that causes SYNIND to work is PRED forming a thought that predicates a property *F-now* of an individual *a* (such as *Fido is ill-tempered now*)
- It then issues an order for a synergic association to be effected between the concept of *a* and the concept of F (ILL-TEMPERED, in our example).
- As with SYN, the strength of the association will depend on the amount of information available.

Again, for SYNIND to be able to create contentful states there needs to be causal grounds for the fact that cues of the type (PRED's creating a thought *Fa now*) go together in a sufficient number of times with the fact that *a* has a certain probability of being F again in the future. The content of the synergic association created by SYNIND is, then, *There are causal grounds which explain the fact that there is a certain probability of* a *being* F *in the future.*

We are interested merely in the general features that make something a modal thought about an individual and, to this effect, characterising a system such as SYNIND which produces just this kind of thoughts is enough. The particular way in which SYNIND does this is irrelevant for the purposes at hand. One such way may involve making the possessor of the concept of *a* go into a state similar to Empedocles's PREPARATION whenever it forms a thought with the content a *is around*. More sophisticated cognitive systems will have a very context-dependent implementation which varies whith every *a*-involving thought the cogniser may entertain. On the other hand, SYNIND is but one example of a mechanism that produces modal contents about individuals. Other mechanisms may use other entirely different cues, apart from a judgement of *a* being *F* on some particular occasion: testimony, for example.

16 In fact, although I have not stressed it for ease of exposition, the same happens with kinds: in the inference from *Fido is ill-tempered* to *Dogs may be ill-tempered* we do not simply reach a logical consequence of the premise, one, for instance, that would be true if Fido is the only member of the species that may be and is ill-tempered. In the kinds-inference too, the conclusion should be read as pointing to a feature of the causal structure of doghood that makes ill-temperedness recur with some probability along the species.

What SYNIND does is creating thoughts about still-open possible courses of the actual world; it encodes the information that, *e. g.*, Fido is such that the future is open with respect to its being ill-tempered again. More sophisticated producers may produce other, more sophisticated contents about still-open possible courses of the actual world. In summary, it does not seem far-fetched to suppose that there is a naturalistically-acceptable reduction along these lines of contents involving what we may call the *in situ possible*:

IN SITU POSSIBILITY: It is *in situ possible that p* at some time *t* ($\Diamond^{t}_{i.s.} p$) iff there are causal grounds for the fact that the probability of *p* after *t* is higher than a threshold *T*[17].

For example, it is in situ possible today that tomorrow will rain iff the world is such that certain causal processes constitute a probabilistic propensity to rain that underlies a high enough probability of tomorrow raining. It will cease to be in situ possible the day after tomorrow, if tomorrow does not rain. Mental contents involving in situ possibilities refer to these causal processes[18].

I should insist once more that, maybe in despite of appearances, our ability to entertain thoughts about in situ possibilities is not constitutively linked with whatever might be useful to us -so that thoughts involving useless in situ possibilities go unexplained by the present account. Instead, mechanisms such as SYNIND tap into causally-grounded correlations -say, between a cue and the fact that a certain probabilistic propensity exists- which may be in place far beyond what we find useful -*e. g.*, outside our light cone. If one of these modal-thought producers has been produced by another mechanism, and not directly selected for, its usefulness may be exactly zero.

5.6.1 *Everyday Possibility*

In situ possibility is a good candidate for the kind of modal contents that may have naturally emerged in the first place. They help cope with what is yet to come, and do so in a cost-effective way. Also, it provides what appears to be a basic building block for our full-blown modal competence. I wish now to indulge in an educated guess as regards how the process to such competence may go[19].

17 The threshold is there to ensure that in situ possibility remains an all-or-nothing affair. Alternative constructions -in which something is in situ possible to a higher or lower degree- are obviously possible.

18 Assuming that probabilities are meaningless without a partition of the space of possible states of affairs, one natural question regarding this definition is: the probability of *p* with respect to what? Well, if we remember the kind of causal mechanisms we have proposed as underlying our modal contents, synergic associations mediate a nuanced response between two different possible kinds of states of affairs. In the three-frogs example, the two possible kinds of states of affairs were *There being a badfly around* and *There being a goodfly around*. In that situation, the probability of each of these states of affairs is to be calculated against a partition that includes both states of affairs and, maybe, a *Things are otherwise* state of affairs to cover all other cases. SYNIND has been selected for enabling these kinds of nuanced responses; it is to be expected that examination of the causal grounds that have facilitated this selection -such as the causal grounds of the relative frequency of goodflies in our example- will lead us to the right partition against which we should calculate the probability of *p*.

19 The following subsections may be seen as elaborations upon Graeme Forbes's "branching conception of possible worlds" Forbes (1985, p. 148f), which, in its turn, is an elaboration of Kripke (1980)'s terse footnote 57, p. 115. See also Pérez-Otero (1997), Mackie (1998), Mackie (2006).
The picture of the modal realm that emerges from the discussion to follow, nevertheless, is a (smallish) proper subset of Forbes's.

The second step on the way to such competence may be a temporal quantification over in situ possibilities:

EVERYDAY POSSIBILITY: It is everyday possible at t that p ($\Diamond_{ed} p$) iff, for a certain contextually relevant time $t' \leqslant t$, it is true that $\Diamond^{t'}_{i.s.} p$.

Everyday possibilities seem to be involved in most of the modal contents we entertain outside the Philosophy department. Take, for example,

> The *James Caird* could have missed South Georgia[20].

It is natural to read this statement as saying that there was a contextually-relevant time (*e. g.*, shortly after the *James Caird* left Elephant Island) in which it was in situ possible that it missed South Georgia; *i. e.*, shortly after this lifeboat abandoned Elephant Island, the probability of its not reaching its destination (as opposed to simply getting lost in the frozen sea) was suitably high.

5.6.2 *Everyday Counterfactual Conditionals*

It appears that judgements involving everyday possibility are what Van Inwagen (1998, p. 73) claims we can know using our 'ordinary human powers of "modalization"'. I will discuss epistemology in the next chapter, but here way can already see that most everyday modal contents seem to be informatively paraphrasable in terms of everyday possibilities. Take one of Van Inwagen's examples of everyday modal truths we can know:

> We'd have had more room if we'd moved the table up against the wall (*ibid*.)

The natural extension of the idea of everyday possibilities to counterfactual conditionals such as this one is to suggest that such conditionals point to the following fact: the antecedent is everyday possible -that is, there is a certain contextually-relevant time such that, at that time, there is a high enough probability of the antecedent holding- and, at that very same contextually relevant time, the indicative conditional linking antecedent and consequent is true. That is, the trick is letting the antecedent fix the contextually relevant time and then use that time to evaluate the conditional. Something along these lines:

EVERYDAY COUNTERFACTUAL CONDITIONAL: $p \ \boxdot\!\!\!\rightarrow_{ed} \ q$ iff for a certain contextually relevant time $t' \leqslant t$, $\Diamond^{t'}_{i.s.} p \wedge (p \rightarrow_i q)$.

That is, we would have had more room if we had moved the table up against the wall if and only if it is everyday possible that we had moved the table up against the wall, and, at that time, if we move the table up against the wall we have more room. This *indicative conditional* $\rightarrow_i$ is not the mere material conditional; roughly, it records a relation between p and q that is analogous to the relation between *F is around* and *G is around* in the disjunctive recruitments discussed in 3.1: there are causal grounds for the fact that, if p, then q to a high enough probability.

It may be suggested that we do not need the antecedent to be everyday possible. After all, the consensus is that counterfactual conditionals

20 If you don't know what this is all about, you definitely need to read about Shackleton's expedition to the South Pole. For example, Lansing (1999).

with impossible antecedents are vacuously true[21]. Asking for the antecedent to be possible, on the other hand, helps fix the right time at which the conditional has to be evaluated. Besides, everyday counterfactual conditionals -such as Van Inwagen's example- are always such that the antecedent is possible. They are popular because they let us think about what would have happened if one of the open possibilities in the past had become actual.

It is likely that our everyday abilities to modalise are restricted to entertaining everyday possibilities and everyday counterfactual conditionals. We may, nevertheless, easily enlarge the class of naturalisable modal contents with a couple of simple modifications to the definitions just introduced.

5.6.3 *More Metaphysical Possibilities*

We may, first relax IN SITU POSSIBILITY, allowing that something is an open future possibility if its probability is nonzero.

OPEN POSSIBILITY: It is *an open possibility that p* at some time *t* ($\Diamond^t_o p$) iff there are causal grounds for the fact that the probability of *p* after *t* is nonzero.

Now we can backtrack the open future to any point in time -not just a particular, contextually-relevant relevant one. If it is true at some time that *p* is an open future possibility, *p* is metaphysically possible.

METAPHYSICAL POSSIBILITY: It is *metaphysically possible that p* ($\Diamond p$) if $\exists t \Diamond^t_o p$.

METAPHYSICAL POSSIBILITY covers a very sizeable chunk of the space of possibilities[22]: *This table might break*, *I might have been born in November*, *Napoleon might have won at Waterloo*... And we can naturalise contents involving such metaphysical possibilities because we can naturalise their ingredients: first, we can naturalise contents involving future contingents -I have devoted the first half of this chapter to give reasons to think that this is so-; and, second, we can naturalise contents about the past -I am simply assuming without any argument beyond the existence of the general content-naturalisation strategies developed in the first chapters of this work, which do seem to apply straightforwardly to past events[23]. We may also assume $\Box p \leftrightarrow \neg\Diamond\neg p$[24]. See figure 15.

The contention is not that METAPHYSICAL POSSIBILITY holds a priori. It need not. For it to hold a priori, thoughts involving metaphysical possibilities would need to be formed out of the concept of past times, and of future contingents. But the producer of metaphysical-possibility

21 For a recent opinion otherwise see Sauchelli (forthcoming).

22 One may even go on to suggest that the space of metaphysical possibilities *is* the space of open future possibilities at all times: even cases that are apparently not covered by METAPHYSICAL POSSIBILITY, in fact are: *Bachelors cannot be married* would, if so, mean that there are no times in which the probability of a married bachelor is nonzero. Nevertheless, it looks as if the truth-maker of this statement has nothing to do with the branching time, and this may point to the fact that there are further metaphysical possibilities that the ones covered here; this is why I'm refraining from writing "and only if" in METAPHYSICAL POSSIBILITY.

23 For the naturalisation of *There was a fly here 30 seconds ago* see 4.2.

24 A naturalisation of negation is needed in the long run if we are to use this equivalence, and the naturalisation of negation is notoriously difficult -see the seldom-read last chapters of Millikan (1984) for a teleosemantic first stab. I hope to pursue this issue in future work.

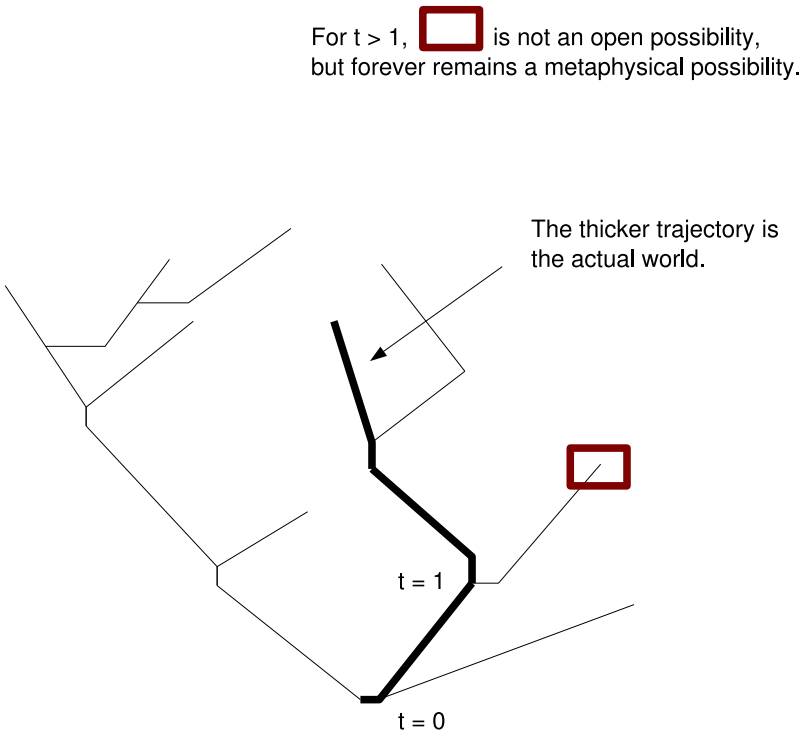

Figure 15: Branching Possible Worlds

thoughts, call it MET, may have been selected for independently, or may have been produced by a selected-for mechanism independently, of those other concepts.

As regards the strict counterfactual conditional, we can maybe suggest that it holds if at the latest past time at which the antecedent is open, the indicative conditional linking antecedent and consequent. The latest time is here playing the role that the class of closest possible worlds where the antecedent is true plays in Lewis's account: worlds that are closest in time to ours are also closest in Lewis's intuitive metric of proximity[25][26].

This, again, makes it neccessary that the antecedent be an open possibility at some time or other, and thus metaphysically possible. If

25 This is just a subset of the picture of counterfactuals that Lewis has in mind -one in which we only recognise possible worlds that are trajectories of the actual world which actualise different options of the open future at different times. See Lewis (1979). Other contemporary accounts of counterfactuals, as far as I can see, overlap with my account in this region. See, *e. g.*, Kment (2006), Williamson (2008).
In particular, Williamson talks of

> the widespread picture of the semantic evaluation of those conditionals as "rolling back" history to shortly before the time of the antecedent, modifying its course by stipulating the truth of the antecedent and then rolling history forward again according to patterns of development as close as possible to the normal ones to test the truth of the consequent. Williamson (2008, p. 150)

The semantics of counterfactual conditionals sketched in this chapter explains why such a method is reliable.

26 There is empirical data to sustain this! Variation of physical magnitudes over time seems to follow power laws, sometimes very closely. See Bell (2008, p. 53) and references cited therein.

we wish to include the case of impossible antecedents, we may use a definition by cases:

COUNTERFACTUAL CONDITIONAL: $p \boxdot\!\!\rightarrow q$ if

1. $\neg\Diamond p$, or
2. $\exists t \left[\Diamond_o^t p \wedge \forall t' \left[\left(\Diamond_o^{t'} p\right) \rightarrow t \geqslant t'\right] \wedge (p \rightarrow_i q)\right]$.

5.6.4 *Modal Contents Naturalised*

To sum up, we have now before us a general strategy for the naturalisation of modal contents: the reference of $\Diamond p$ is a complicated state of affairs, a quantification over probabilities and times. None of the components of this state of affairs are such that contents involving them are suspicious of being impossible to naturalise. Contents involving the past are clearly naturalisable along the lines of the etiosemantic theory I have developed in the first chapters. Contents involving probabilities, indeed, are more suspicious, but in this case we have seen a proposal in some detail: the mechanisms we have called *synergic associations* allow us to see how contents involving objective probabilities may be naturalised. To be sure, much detail remains to be provided -importantly, a naturalised account of contents involving quantification and negation should be developed- but we now have more solid grounds to believe the task to be possible.

There is a further, lesser problem to be recorded here. We have seen how contents involving probabilities may be naturalised, and we have expressed the hope that contents involving the past may also be naturalised. So, at least the following route to the naturalisation of modal contents is available: we entertain them by using our concepts of the past and the probable. The problem with this route is that the relation between these concepts and the concept of the possible does not seem to be a priori, as it should be if we really thought about the latter using our concepts of the former. There are two options here: we may dig our heels, and defend that, indeed, the relation between such concepts *is* a priori. This may be close to what Kripke and Forbes believe. The other option is to defend that the relation between such concepts is fully a posteriori, and then we need a story that explains the existence of concepts that is enabled by the existence of the complicated states of affairs that, we have said, are modal states of affairs. This story either exists or it does not, and it may be epistemically impossible for us to discover it. None of the options, therefore, is entirely out of the question. It is very difficult to ascertain whether the relation between our concepts is, or not, apriori, so it is difficult to know whether the first option is open at all. It may be impossible to discover the facts that have made an independent concept of the modal realm useful, so it may be impossible to know whether the second option is an option at all.

In the following section, to further clarify the view, I will draw a number of consequences of the branching metaphysics of modality I have quickly sketched in these last sections.

5.7 SOME CONSEQUENCES

5.7.1 *The Right Modal Logic*

According to the semantics of modal operators sketched above, the axiom

M: $\Box A \rightarrow A$

is met. If there is no moment in time in which $\neg A$ is an open possibility, it is true in particular that the actual moment is such that $\neg A$ is not an open possibility -let alone true.

Also,

4: $\Box A \rightarrow \Box\Box A$

holds. If there is no moment in time in which $\neg A$ is an open possibility, then there is no moment in time in which it is an open possibility that there is a moment in time in which $\neg A$ is an open possibility. I am assuming here -and everywhere- that the fact that a possibility becomes actualised at any time can have no impact on which possibilities were open at prior times. Otherwise put,

ETERNITY: The facts about which possibilities are open at any time time are eternal facts, with beginning but no end.

This is the most natural view of time, but by no means the only available, and other options will maybe allow rejection of 4. I cannot defend my choice here, but will simply record this as an additional committment of the kind of naturalistic account of modality I am developing here. Finally, the axiom characteristic of S5:

5: $\Diamond A \rightarrow \Box\Diamond A$

Appears to be met as well, once we accept ETERNITY: if there is a time at which A is an open possibility then there is no time at which there is no time at which A is an open possibility.

According to this understanding of modality, then, we have the nice result that the right modal logic is S5.

5.7.2 *Determinism, Counternomic Possibilities*

It should be noticed that according to the account of modality just sketched, the following possibility does not exist, or at least is not covered by the account: determinism is true -that is, roughly, for every proposition of the type "at t, p" only it or its negation is an open possibility at any time- but, nevertheless, things could have been otherwise. That is: metaphysical possibilities are constituted, in the way described above, by facts about the open future at different times.

It is likely that many scientists of the Enlightment were determinists, and it is also likely that some of them believed that, even so, things could have, in some sense, been otherwise. Perhaps the idea was that the world's timeline was non-branching but that, alongside, there were other possible timelines. All of this is out of the question if the kind of branching account of modality I have presented is correct.

There is another, perhaps more worrying counterintuitive consequence of this picture of modality. Under the plausible assumption that

natural laws are not fixed at any particular moment in time, they are also metaphysically necessary. Although not the mainstream position, we are used to necessitarianism about natural laws (*e. g.*, Shoemaker (1998)) but the branching conception of modality offers different reasons than the usual Kripean considerations about the semantics of terms such as "mass". It also has a bonus (malus, really) counterintuitive consequence in the vicinity: suppose there is an initial moment of time, t_0. All intrinsic facts about t_0 -not involving later times, that is- are necessary: the number and distribution of particles, their temperature, etc. could not have been otherwise. I expect many readers to protest at this point. They surely know some of these states of affairs to be contingent? Against this retort I can only say the following: the reader's state of conviction might be illusory. And there is a powerful reason to accept this kind of metaphysics of modality: it allows a substantial naturalistic explanation of our epistemic access to modal facts. So, I will ask the reader to bear with me and, at the end of this work, ponder whether the weight of this (admittedly uncomfortable) result is or not compensated by the advantages that accrue from accepting the theory.

5.7.3 *Necessity of Origin*

Finally, I would like to sketch a way in which this branching conception of modality makes essentialist theses such as the necessity of origin come out true.

There is a more optimistic way to read the uncomfortable result that the state of the world at t_0 is necessary: this would be a particular instance of the widely accepted principle that some properties of the origin of entities are essential of those very entities. Thus, the egg and sperm that originated the zygote that in turn became me are essential to me: I could not have originated from other egg and sperm. Also, the material, and maybe the process of construction, which are used in building a certain table are commonly regarded as essential to that very table. In the same vein, maybe it is essential to our world being what it is that it has the properties at t_0 that it does.

I do not think much of this redescription of necessitarianism about t_0. Saying that its intrinsic properties at t_0 are essential to our world seems to convey the idea that whatever world it is that does not have said properties at t_0 is not our world. But no world fails to have the actual world's properties at t_0. The fact that our world has these properties essentially is, at best, trivially true. Anyway, this friendly attempt at redescription points out to a possible strategy to defend essentialist principles such as those described above.

The simplistic way to implement the strategy (more or less endorsed by Forbes (1985), and criticised by Pérez-Otero (1997) and Mackie (1998) on the counts I summarise below) would be as follows: first, assume that time is forward-branching only, *i. e.*, time only branches into the future, but not into the past; second, assume that possible worlds that contain a certain entity a must be identical to the actual world up until the time in which a comes into existence. That is, every a-containing possible world shares an initial segment with the actual world that includes a's coming into existence. If so, a's origin is, indeed, necessary.

This defense of the necessity of origin comes at too high a price: it makes every event happening before a's time, up to and including its origin, essential to a. So (to use an example by Pérez-Otero) if Kant

sneezed n times during his lifetime, there is no possible world in which he sneezed n+1 times and I exist. This is unacceptable.

One alternative to this is proposing, with Mackie, that the way to think of de re possibilities is one

> in terms of divergence *into the future* from 'the actual course of events', even if we are not very strict or precise about the extent of match that is involved when we speak of 'the actual course of events'. Mackie (2006, p. 103)

The idea, I take it, is having a notion of the actual course of events that counts, for the purposes of evaluating de re possibilities about me, as the same courses of events those that only differ in irrelevant respects such as the number of sneezes that afflicted Kant during his lifetime. The problem, then, is to provide a principled way for grouping together irrelevantly-different courses of events.

We should remember from earlier chapters that individuals are hosts of causal mechanisms keeping together clusters of properties. Individuals, like kinds, are things about which we can learn in some circumstances things that remain true in many other circumstances. It is this feature that has enabled the existence of individual-involving contents and, eventually, individual concepts.

Keeping this in mind, one natural suggestion regarding the individuation principle for courses of events that underlies evaluation of de re possibilities regarding an individual a may be one of causal independence between the facts that differ among said courses of events and the causal mechanisms that keep a together, so that:

CAUSAL ISOLATION: A course of events is one in which a exists if and only if

- it overlaps with the actual course of events up to and including the time of a's origin. Or
- the events in which it differs from the actual course of events do not prevent the causal mechanisms constitutive of a from existing.

It seems to be the case that the causal influence of Kant's sneezing just one extra time would have not been amplified, from the late 18th century to the late 20th century and from Königsberg to Spain, in a way sufficient to prevent the sperm and egg which originated me to form the zygotic proto-me. This is why, according to CAUSAL ISOLATION, a world that differs from our own only in that Kant sneezed one more time is a world in which I exist[27]. Of course, if, unexpectedly, Kant's sneezing *has* an impact in my parents' sperm and egg then Kant's sneezing n times is necessary to my existence[28].

There is a substantial question here of why count the union of sperm and egg as the debut of the casual mechanisms constitutive of a. It is very natural to say that a couple's plan to have a baby is also a good candidate to being part of these causal mechanisms. Why isn't it?

27 What counts as the same egg and sperm in a world which diverged from our own before they came into existence? Reapply CAUSAL ISOLATION for said egg and said sperm.

28 An important class of events that meet CAUSAL ISOLATION are those which are not in a time-like relation. That is, if from the point in spacetime in which Kant sneezed it is impossible to reach the point in spacetime in which I am conceived at the speed of light, then Kant's sneezing has no impact in my existing. But there are surely other, less clear cases.

The view that I have been advancing has it (simplifying a lot) that individual-involving contents (and, eventually, singular concepts) emerged because they were useful in keeping track of individuals. Individuals keep their properties across time so that what we learn upon one encounter with them carries over to other encounters.

Now, even if there are some individual's properties the recurrence of which may be explained by a couple's intention to have a baby -e. g., the baby's family resemblance with their parents-, the truth is that the kinds of mechanisms that are kicked off by the formation of the zygote are much better at keeping some individual's properties together -the general looks, the sex, a general propensity to be melancholy, say- and their efficiency does not wane until the death of the individual[29]. This is maybe why our concepts of individual human beings are attuned to the coming into being of the mechanisms that the existence of the zygote facilitates. That they are so attuned or not is, on the other hand, a purely empirical question.

All in all, the general explanation of our inclination to count a possible course of events as one in which *a* exists is that whatever happens in this alternative event should not affect what *a* is. And what *a* is, in turn, is fixed by what individuals in general are; that is, by the kind of structures that have enabled the selection for a producer of individual concepts.

5.8 EPISTEMIC POSSIBILITY

It is not totally unlikely that the reader has intuitively identified the kinds of probabilities we have been talking about with *creedences*. In the three frogs example, there is a natural way to describe Empedocles situation which is: *for all Empedocles knows*, that fly (the fly approaching) has a 30% change of being a goodfly, and this is the condition that he is adapting his behaviour to. This is so even if the fly approaching is a badfly (and let us assume that if something is a badfly then it is metaphysically impossible for it to be a goodfly.) Thus, the state formed by the synergic association of Empedocles's concepts FLY and GOODFLY[30] would be no less and no more than Empedocles's placing a 30% creedence on each fly being a goodfly.

This way of looking at matters is incorrect. For all we have said, Empedocles may only have: a mechanism whose positives have the content *There is a fly around* (the Fly-mechanism); a mechanism whose positives have the content *There is a goodfly around* (the Goodfly-mechanism); and the synergic association between them that, I have argued, has the content *There is a causally-grounded relation between flies and goodflies such that* $\mathtt{P(Goodfly|Fly)} = 0,3$. Empedocles has no state whose content involves an *individual* fly; no state with the content *That fly is a goodfly* or *Fido is a goodfly*. He is, then, unable to predicate of any individual fly that it might, or might not, be a goodfly. The only relation he represents is one among natural kinds -flyhood and goodflyhood- and this is a *bona fide* probabilistic relation, not merely an epistemic one: flies *do* have

29 Or maybe shortly after; this is maybe why, although there seems to be a clear point in which the individual starts to exist, the point at which it ceases to exist is less clear.

30 Strictly speaking, Empedocles does not have these concepts. Only unarticulated thoughts with the content *There is a fly around* and *There is a goodfly around*. Talk of concepts is more natural, and we already know how to account for synergic connections at this level, so I'm helping myself to the loose talk.

a 30% chance of being goodflies, even if for each individual fly either it is a goodfly or it is not.

It is true that the way in which we have rendered the content of the positives of Empedocles's Fly-mechanism (*There is a fly around*) makes an epistemic reading almost irresistible: a fly approaches and the *There is a fly around* state lights up in Empedocles's brain. Given that he has a standing belief with the content *Flies have a 30% chance of being goodflies*, isn't he jumping to the conclusion that *the fly approaching* has a 30% chance of being a goodfly? Isn't that what his investing in *mg* amounts to?

It has happened in earlier chapters: the natural rendering of the content of Empedocles's states is somewhat misleading. Quine-style alternative renderings such as "Lo! More fly!" do not lead us to this mistake: Empedocles is detecting "more fly" and, seen as an uncountable, every "portion of fly" has the same probabilistic relation with goodflyhood. Empedocles's states encode genuine knowledge about the world -the evolutive history of the states playing the justificatory role. It is simply not knowledge about individual flies, but about the relation between flyhood and goodflyhood.

5.8.1 *The emergence of epistemic modality.*

In chapter 3, I have developed the materials to account for some individual-involving contentful states, and later, in chapter 4 I have provided an interesting set of sufficient conditions for the presence of concepts of individuals. If we endow Empedocles with the ability to entertain individual-involving thoughts, we may start to see states with epistemic-modal content.

Suppose now that Empedocles is able to entertain the thought *That fly is around* whenever he detects a fly around him. More should be said, of course, to motivate just how and just why would Empedocles go about doing this, but I hope that, remembering the discussion in chapters 3 and 4, we can grant the point for the sake of discussion.

It may be turn out useful to implement a top-down process of converting predications of possibility at the kind level into predications of posibility at the individual level: whenever the agent finds out that Fs might be Gs, it effects a synergic association between every individual concept of an F, and the concept of G.

We may want to say that this more sophisticated version of Empedocles is entertaining thoughts along the lines of *That fly might be a goodfly*. For instance, we may not want to say that Empedocles is wrong in thinking this, even if it is metaphysically impossible that a badfly had been a goodfly, and if "that fly" is a badfly. It should be noted that *That fly might be a goodfly* is simply an intuitively appealing content-attribution for one of Empedocles's mental states, but not a content attribution that follows the recipes I have been defending throughout this work. Those recipes build contents out of the explanations of the existence of the contentful state and, in this case, the only explanations available are those subserving contents attributions of probabilistic relations among the fly and goodfly kinds, and of the presence of an individual fly. There are no extra causally-grounded correlations in the external world to be involved in the content of this mental state. And the causal grounds that are available, as we have seen, only warrant the

content attributions: "$P(\texttt{goodfly}|\texttt{fly}) = x$" on the one hand, and *That fly is around* on the other.

That is, we may advance the hypothesis that the content we would express as *That fly might be a goodfly* with an epistemic *might* is the same content as $P(\texttt{goodfly}|\texttt{fly}) = x$ *and that fly is around*.

If we are in a daring mood, we may take these considerations as evidence for an analogous paraphrasis of our full-blown epistemic-modal talk:

EPISTEMIC POSSIBILITY: It is permissible for S to find Fa epistemically possible if[31] S (permissibly) believes that

1. $P(F|G) = x$ for a relevant property G and a sufficiently high x, and
2. Ga

Where the "sufficiently" and the "relevant" in 1 depend on the context of evaluation. Thus, for instance, *This table might be Swedish*, with an epistemic *might*, means that the table belongs in a kind (e .g., smart-looking, clear-wood tables) such that the probability of such tables being Swedish is high enough in the context of utterance -maybe a casual conversation among friends, where the issue is not really critical for anybody-. And *This might be the winning lottery-ticket* means that the ticket belongs in a kind (*e. g.*, valid lottery tickets) such that the probability of such tickets winning is high enough in the context of evaluation -different values will be operative in a conversation between two burglars that have broken into the house of a lottery-winner, and in a TV ad promoting a lottery. I cannot go into a full discussion of epistemic modals here, and simply advance these considerations as directions for future work.

5.9 HUMEANISM ABOUT PROBABILITIES

I want to put an end to this chapter considering the following broadly Humean objection:

> For every member of Empedocles's family line, the existence of the synergic association of its M and PREPARATION mechanisms may be explained by appealing only to the statistical frequency of goodflies in the fly population *in the past*. No mention of objective probabilities is needed. The appeal to them does no real work. With suitable modification of your content-attributing recipes we could get to an attribution along the lines of *In the past, the proportion of flies that were goodflies was thus and so*. This is ontologically more circunspect, and is therefore to be preferred.

Let me for a moment suppose that this, more circunspect content attribution is to be preferred. The first thing to say is that we should really insist in the need to examine the theory that yields it. Maybe that theory gives all sorts of counterintuitive content-attributions; then again, maybe it gives better and more elegant results that the theory I'm advocating for. The point is, we don't know, and theories should be evaluated as wholes. It is in general too easy a criticism to point at a single counterintuitive result, without providing arguments to the effect

31 Not "... and only if", of course.

that the result in question could be avoided without loosing overall plausibility at other corners of the theoretical building.

Having said that, we cannot dispense with objective probabilities, even if we wished to do so. Of course, in a sense, the existence of Empedocles's synergic-association state is explained by the past history of statistical frequencies between flies and goodflies. Even so, in order to give a common explanation for the survival of PREPARATION in Empedocles's family line, you need to get to the level at which homeostasis takes place. What the Humean is not taking in consideration is the need for having all (or at least most*) of Empedocles's family members intertwined in the same explanation with the probabilistic relation between fly and goodfly. This you can't get just by appealing to the statistical frequencies in the past of each member; that each member happens to stumble upon the same frequencies would be left unexplained. This would be a case of sheer luck, giving rise to no contentful state -or so I argued in chapters 1 and 2.

So, at least this can be said: if the Humean is right, and frequencies must be left unexplained, then Empedocles doesn't have contentful states that refer to probabilities after all -nor do they refer to statistical frequencies or any cognate thereof.

6 MODAL EPISTEMOLOGY

This final chapter is dedicated to advancing my final proposal for the naturalisation of modal epistemology. In many respects, the account to follow simply falls out from what I have said in the chapters leading to this point. The one new idea introduced in this chapter is a proposal for a solution to the Generality Problem in reliabilism (6.1.1). From this solution, a particular brand of reliabilism I shall call *ecological reliabilism* emerges. I will suggest that modal epistemology may be simply a special case in ecological reliabilism (6.2).

At this point we may ask, do we really need a reliabilist modal epistemology, even if such a thing is workable? There has been an enormous amount of high-quality work in neo-rationalist accounts of modal epistemology; what is the problem with that? In 6.2.1 I will elaborate on one problem with conceivabilism, a prominent example of neo-rationalist modal epistemology: conceivabilists have devoted their efforts at dispelling or otherwise avoiding several famous counterexamples to the rationalist main insight -in one version at least- that conceivability entails possibility. The account we are left with, once the counterexamples are warded off, is to be accepted simply on the grounds of its lack of counterexamples. I will defend that this is unsatisfactory, at least in the sense that any other account which provides a more substantial explanation of the link between conceivability and possibility is to be preferred. I will, then, suggest that the package deal provided by etiosemantics and ecological reliabilism, together with the branching conception of modality, may provide such a substantial explanation of our modalising powers. It is, therefore, to be preferred.

6.1 CONTENT AND EPISTEMOLOGY

In accounts of the kind I have been advocating, there is a quite straightforward relation between the conditions for a state to have the content it does and some conditions which have been defended as underlying epistemic properties such as *Being justified*, or *Constituting knowledge*. Take, again, the simplest content-attributing conditions I discussed in chapter 1 for a content of the type *There is an F around*. I defended that the state consisting of a certain mental mechanism M's being *on* has that content if M was selected for indicating any of a number of properties, and this selection was enabled by the fact that all or many of these properties belong in a Homeostatic Property Cluster: a cluster of properties such that there are causal grounds for the fact that they co-recur frequently together. As a useful abbreviation, let me introduce the notion of sufficiently reliable:

SUFFICIENTLY RELIABLE: A state S is a sufficiently reliable indicator of the existence of fact F according to threshold value T if and only if: $\alpha(S, F) > T$.

Where $\alpha(S, F)$ is the correlation coefficient between S (say, M's being *on*) obtaining and F obtaining[1]. The relevant probability space (i. e.,

1 More strictly, the correlation between their indicator random variables.

which possible or actual situations should be taken into account when calculating the probabilities used to calculate $\alpha(S, F)$) depends on the context -*cf*. Williamson (2000, p. 84).

We can now note that M's being *on* is a sufficiently reliable indicator of the fact that F is around, where the relevant probability space is fixed by the context in which the homeostatic mechanism that sustains F-hood is operative -*cf.* 1.4.5 for details. That is, a certain property (*Being a black speck*, say) causes M to go *on*, and instantiations of this property, in turn, covary positively with the presence of F-hood, in a certain spatio-temporal context, which is the one that fixes $\alpha(S, F)$ -say, ponds where frogs endowed with M have normally lived, and which they have shared with flies. Here, the threshold value T is, simply, the value afforded by the causally-grounded correlation between instantiaton of *Being a black speck* and the presence of flies. This value has helped M thrive and get fixated in the frog population. Maybe a lower T would have been just as good for selection, but this is irrelevant.

Following the terminology introduced in chapter 3, we may say that M follows the Procedure: *Fire now if a black speck is around now*[2]. Also remember from 3.6 that every Procedure comes with an ecologically-fixed context. Now, the following paraphrases of epistemic idioms are not totally implausible:

- M's going *on* is *justified* only if M has followed its Procedure in its ecologically-fixed context -see 3.6 for more on these notions. Derivatively, we may say that M's possessor is justified in judging that an F is around in this situation.

- M's going *on* constitutes *knowledge* if the HPC that explains M's reliability in the ecologically-fixed context is responsible both of M's going *on* and of F being around. Derivatively, we may say that M's possessor knows that an F is around in this situation.

Where *an HPC being responsible* of M's going on involves the HPC occupying the same position in the causal network leading to M's going *on* that it has occupied in a sufficient number of occasions of causing M to go *on* in the past. On the other hand, an HPC being responsible of F being around may simply involve partly *constituting* the fact that F is around. So, the HPC in question constitutes flyhood, which is why it is, in this latter sense, responsible of the presence of a fly around. When a fly causes M to go *on* in the usual way -that is, by the instantiation of *Being a black speck* that belongs in the cluster of the HPC *fly* causing M to go *on*-, this mental state constitutes knowledge that a fly is around. If, on occasion, a random black speck causes M to go on in the ecologically-fixed context, M's being *on* is justified but does not constitute knowledge. If, furthermore, a fly *is* around -but it is, say, an albino fly and has not interacted with M-, M's being *on* is justified, true but does not constitute knowledge.

For another example, consider the American lobster I discussed in 3.9. A lobster L_i has a mechanism M_{ij} whose positives have the

2 We may want to make the assumption that M's going on corresponds, at the "personal" level, with M's possessor's (proto-)judging that an F is around -I said something in favour of this assumption in 3.1.2-, but we should be careful: this does not mean that M's following the Procedure above corresponds to M's possessor following the personal-level Procedure *Judge that a fly is around whenever a black speck is around*. The fact that Maurice is caused to judge that a fly is around by a black speck being around does not mean that he can entertain that Procedure -which, I take it, is need for following it at the personal level. He may be unable to entertain contents involving black specks, for example.

content *Lobster L_j is around*. M_{ij} follows the Procedure *Go on when chemical compound UCS_j is in the water*. This Procedure comes with an ecologically-fixed context which includes part of the Atlantic coast of North America -and which does not include tanks of salt water in ethology departments. According to the paraphrases above, when M_{ij} goes on in the presence of chemical compound UCS_j somewhere in the coast of Maine it does so justifiedly. If, moreover, it was lobster L_j that released UCS_j, it constitutes knowledge.

6.1.1 *Ecological Reliabilism*

These paraphrases, which I will be presently stating in a somewhat more formal fashion, may help negotiate several difficulties with other broadly reliabilist proposals. Take one popular theory about justification in contemporary epistemology: *process reliabilism* (first presented in Goldman (1976/2000), see also Goldman (1986) and Goldman (2008)). A simplified version of the main thesis of process reliabilism is:

RJ: A belief is justified if and only if it is produced by a process that reliably leads to true beliefs. Conee and Feldman (1998, p. 1)

Many interesting objections have been advanced against process reliabilism. I will only take up one here, arguably the most influential: the *Generality Problem* (Conee and Feldman (1998)).

The objector asks us to notice that reliability is a property of process *types*. A process token either fails or succeeds in producing a true belief, but it is not *reliable* in any clear sense. Now, any process token belongs to innumerable process types. Take, for example, the case of Smith forming the belief that there is a maple tree nearby on the basis of a perception of the tree in question through a (solid, but transparent) window. This seems a clear case of justified belief, but it is unclear *which* is the process whose reliability supports this intuitive conclusion:

> The token event sequence in our example of seeing the maple tree is an instance of the following types, among others: visually initiated belief-forming process, process of a retinal image of such-and-such specific characteristics leading to a belief that there is a maple tree nearby, process of relying on a leaf shape to form a tree-classifying judgment, perceptual process of classifying by species a tree located behind a solid obstruction, etc. The number of types is unlimited. Conee and Feldman (1998, p. 2)

And some of these process types will be extremely reliable -*e. g.*, relying on leaf shapes-, and some others very unreliable -*e. g.*, relying on perceptions from behind a solid obstruction, which more often than not will be opaque. Without a principled way to choose the relevant process type, process reliabilism is not an assessable theory. Conee and Feldman consider proposals as regards which is the way to choose among competing type-candidates which come from three different approaches:

- *Proposals based in common sense*. The relevant process type should be the one the common sense dictates. This seems to leave out, in the example above, the type *perceptions from behind a solid obstruction*, which common sense quite clearly does not endorse. But,

Conee and Feldman retort, the are many other candidates in the example that common sense is confortable with: visual perception, tree-identifying process, etc. Process indeterminacy follows. I should add that an undischarged appeal to common sense, even if it were not to suffer from the mentioned shortcomings, would also be suspicious for my purposes: we are in the business of sketching a naturalistic epistemology, and the clearly intentional workings of common sense are not a respectable basic ingredient in such an account.

- P*roposals based in science*. There are several:
  - The relevant type for any belief forming process token is the natural kind to which it belongs. Conee and Feldman (1998, p. 10)

  But there are many natural kinds a process may belong to.

  - The relevant type for any process token is the natural psychological kind corresponding to the function that is actually operative in the formation of the belief. Conee and Feldman (1998, p. 11)

  That is, the process that is psychologically real. But there are many processes that are psychologically real in the formation of a belief that *p*. In the example above, the process that goes from the very leaf shape Smith perceives to his belief that there is a maple tree; the process that take any number of perceptions of similar shapes and issues a belief that there is a maple tree nearby, etc. All of them are active in the case in question.

  - The relevant type for any belief forming process token t is the natural kind that includes all and only those tokens sharing with t all the same causally contributory features from the input experience to the resulting belief. Conee and Feldman (1998, p. 14)

  This is too restrictive; beliefs are only considered of the same type if they share all causal antecedents, but we obviously type beliefs more broadly. Another issue one may rise is, how should we type causal antecedents themselves?

  - The relevant type for any belief-forming process token t is the psychological kind that is part of the best psychological explanation of the belief that results from t. Conee and Feldman (1998, p. 17)

  But we should resist the unwarranted assumption that there is *one* best psychological explanation of our beliefs.

- Finally, one may suggest that we do not need a systematic solution to the Generality Problem. For example, one may suggest that it is contextual factors that fix which process is relevant for the attribution in question. But it does not seem to be true that, for most situations in which a belief is formed, there are contextually salient processes. For example, in the perception of the maple tree above, it is not true that there is one, and only one, salient process for the formation of the belief that there is a maple tree nearby.

Conee and Feldmand, therefore, conclude that every proposal fails, and thus, that process reliabilism fails.

I would like to argue that, if what I've said so far in this dissertation is approximately right, Conee and Feldman have overlooked a fourth way of singling out the relevant process type. According to this fourth way, the methodology to come up with the right process type that has produced the token belief B with the content that p would rely on the *processes that endow* B *with the content it has.* We may suggest the following heuristics for finding out the relevant process in the formation of a belief B:

1. Ascertain which are the causal-historical properties in virtue of which B counts as having the content that *p*.

   That is, search through the battery of content-attributing recipes I have been developing throughout the work, and find out which one applies to the belief in question. Let us suppose, for example, that a cue S has caused a belief-producing system BEL to produce B. BEL is able to produce perceptually-based beliefs, which may suppose -simplifying a lot; see chapter 4 for a fuller story- that it belongs in a higher-order reproductively-established family the selection of which has been made possible by the existence of HPC connecting cues such as S correlate with states of affairs such as *p*. In a realistic setting, the workings of BEL will be fantastically complex, and the cue may involve an open-ended amount of retinal shadows.

2. The content-determination will involve a history of selection for the belief in question or, much more likely, a history of selection for BEL, or BEL's producer. The selection in question, if it's to give rise to a contentful state in its product (or product's product, or...) must have been enabled by an higher order HPC -all of this is covered in detail in earlier chapters.

   Now, if there are *n* members in the chain of producers (we may assign number 1 to the selected-for mechanism, number *n* to the contentful state we are interested in) there are *n-1* Procedures, one for each link, which the selection for 1 fixes[3]. Each of these Procedures must be sufficiently reliable in a way that adds up to the fitness-contributing properties of the 1st link. All Procedures have an ecologically-fixed context within which they are fitness-contributing. This context is defined, for each Procedure i, by the ith level of the HPC.

3. Finally, the process relevant to the justification of B is the concatenation of all the n-1 Procedures, together with their ecologically-fixed context. In most cases, a sizeable number of Procedures will be innate, or fixed early in the life of the individual, and only the last few will matter to the justification of the belief. But, strictly, all of them are relevant.

The most general account of justification and knowledge would, therefore, be as follows[4]:

JUSTIFICATION: A belief B with the content that p is justified if and only if

3 See the discussion of a concrete example of the etiosemantic treatment of what I've called ephemeral states -contentful yet not selected-for- in 3.9.

4 A couple of examples of justification and knowledge according to this most general account were already provided at the beginning of this section. Democritus was featured in one, American lobsters in the other.

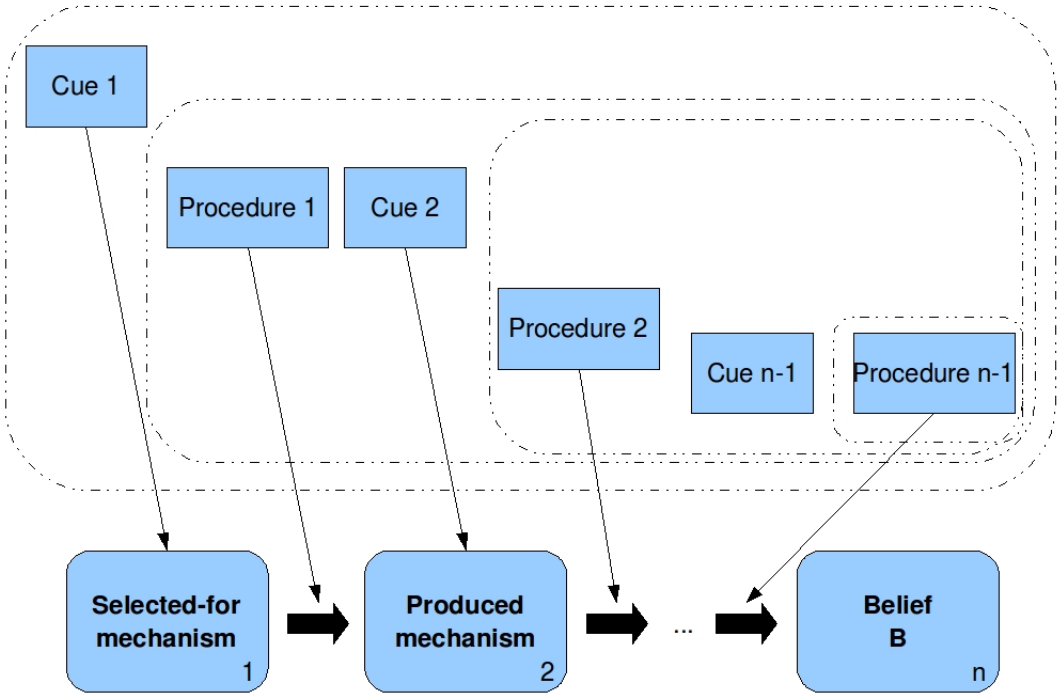

Figure 16: A Concatenation of Procedures and Ecologically-Fixed Contexts.
Each mechanism is produced by the mechanism in the left, the chain bottoming out in the selected-for mechanism 1. The rightmost box stands for a state, not a mechanism: the belief B that *p*.
Each mechanism is caused by a Cue to produce the item in its right, following a Procedure. The fact that Cues and Procedures that are fitness-contributing go together is explained by a set of nested causal grounds (the Nth-order HPC I cite in the main text) which are represented by the dotted-line boxes. E. g., Cue 1 goes together with the fact that Procedure 1 is fitness-contributing -which, in turn, implies the fact that [Cue 2 and the fact that Procedure 2 is fitness-contributing go together, which in turn...].

1. B's content attribution that *p* is the result of a chain of *n* producers, of which the 1st is selected-for.
2. The selection for 1 has been enabled by the existence of an n^{th} order HPC.
3. This selective history, and the n^{th} order HPC, fix a concatenation of Procedures for each link in the chain, together with an ecologically-fixed context for the application of each Procedure. Finally,
4. B has been produced according to the *n* Procedures, such that each was in its ecologically-fixed context (see figure 16).

Derivatively, we may say that B's possessor is justified in believing that *p* in this situation.

We may now provide a general version of sufficient reliability:

SUFFICIENTLY RELIABLE - GENERAL: The process P that gave rise to belief B is sufficiently reliable according to threshold value T if and only if: $\alpha(S, F) > T$. Where *S* is the type *Beliefs produced by* BEL and F is the content of those beliefs.

The process consisting of all the concatenated Procedures that give rise to B is sufficiently reliable in its ecologically-fixed context, with a T fixed by the entirely unpredictable nature of the n^{th} order HPC that B and its chain of producers are tapping into. As I envisage these structures, it may be that each different <Cue, Procedure> pair at each different level *i* gives access to a different HPC at the level *i*-1. Each of which may enable a higher or lower reliability. These are the values of sufficient reliability. They are what they are, regardless of our ability to discover them. So, if I'm right, Goldman gets it wrong when he writes that

> [j]ustification conferring processes are ones with a high truth-ratio. (Just how high is vague, like the concept of justification itself). Goldman (2008)

The degree of reliability of justification-conferring processes is not vague, or not as vague as Goldman assumes. It is simply very well concealed.

We may now, provide the general paraphrasis of knowledge talk according to this ecological variety of reliabilism:

KNOWLEDGE: A belief B with the content that *p* constitutes knowledge if and only if the n^{th} order HPC that explains the sufficient reliability of the concatenation of Procedures in their ecologically-fixed context is responsible both of the existence of B and of the fact that *p*. Derivatively, we may say that B's possessor knows that p in this situation.

The idea, then, is that the principled way to choose the process whose reliability counts in order to ascertain the status of a belief as justified stems from *the process in virtue of which a mental state is endowed with the content it has*. The very nature of content-conferring processes makes them (sufficiently) reliable, and this status is what our epistemic talk tracks: knowledge happens when all goes well, as a result of following the right Procedure in the right context; justification happens when all Procedures are followed, in their right context.

It should be noted that what can be known according to this account far outstrips what has been useful to us or our ancestors. What defines the ecologically-fixed context for each level is the HPC that keeps together a certain type of Cue with the fact that a certain type of Procedure is fitness-conducive -in the simplest cases; see 3.6. The HPC in question may stretch way beyond the instances that have been, will be or may be used by humans and their ancestors. For example, there is an HPC that connects certain colourful shapes with the presence of orchids. It is the HPC *orchid* itself, which has instantiations of such properties as *Being thus-and-so a colourful shape* in the Cluster. Where and when is such an HPC instantiated is independent of human cogniser interested in orchids. Such an HPC may well extend the ecologically-fixed context of a process that takes in such shapes as Cue and issues a Procedure that creates a mental state with the content *There is an orchid around* to areas that are outside the light cone of every human, past,

present or future. Such will be the case, *e. g.*, if orchids outlive humans on Earth[5].

This is how ecological reliabilism deals with the three criteria that Conee and Feldman lay out for a satisfactory solution to the generality problem:

1. *It must be principled, not concocted ad-hoc for each belief*: my proposal is that, for every belief token, the relevant process is the process in virtue of which a mental state is endowed with the content it has. This is an homogenous, principled condition.

2. *It must make defensible epistemic classifications. That is, it must yield intuitive predictions as regards what is or is not justified*: this is difficult to assess in real, human cases because we have very incomplete information about the processes leading to a contentful state. More transparent process attributions, such as "perception" or "recognition of tree-type from leaf-shape" fare better in this count but fail in responding to the other criteria. In favour of the ecological reliabilist proposal we may say two things: first, it might well turn out -I, for one, believe that it would- that the justification and knowledge predictions achieved by a completed ecological reliabilist theory overlapped significantly with our intuitions. Second, we have seen in the simple cases of Democritus and American lobsters above that they, indeed, seem to correspond with our intuitions.

3. *It must remain true to the spirit of the reliabilist approach. That is, it is the reliability of the process in question that settles the epistemic status under scrutiny*: Ecological reliabilism is extensionally adequate, in that the relevant processes will always be sufficiently reliable in the sense above. It may be suggested that reliability is not in the essence of epistemic notions according to this account, but rather that content-endowing Procedures are. Even so, the proposal at hand explains why justificatory processes are reliable. This is, I think, enough to vindicate Goldman's seminal insight.

6.1.2 *When Things Go Wrong*

To help clarify the theory, let me show how this ecological reliabilism accounts for epistemic error. False beliefs are easy: the content of beliefs are fixed in ways which are independent of how the world is in many respects -the immense majority of beliefs have contents that involve these respects. They will, frequently, be false, even if they have been

5 By the same token, we are perfectly able to entertain contents with truth-makers that lie outside our light cone. This is, I believe, the right kind of answer to "useless content" objections to teleosemantics such as Peacocke (1992, p. 132f). Although, of course, the answer comes from the part of etiosemantics that parts company from strict teleosemantics: the content of a concept or an expression depends primarily on the HPC that has enabled its existence; not on, say, the usefulness in principle for a consumer system. Some HPCs do, while it may be that no usefulness-in-principle does, extend their reach outside the light cone of every sentient being.
The "useless content" objection is probably more powerful against mainstream teleosemantics. For a rejoinder, see Millikan (2006). In this paper Millikan defends that it is compositionality -in the very broad sense she understands this notion- that is going to account for representations of facts outside our light cone. I have expressed my doubts about Millikan's account of compositionality in chapter 4.

formed following the right concatenation of Procedures in the right contexts and are, thus, justified[6].

As regards unjustified true beliefs: there may be beliefs formed by a producer that follows the right Procedure but outside its ecologically-fixed context[7]. This fails to meet JUSTIFICATION. Illusions are a clear example: a belief that a red tie is orange, formed after seeing it under strange illumination is probably formed following the right Procedure, in the wrong context. Unjustified beliefs may be true, if the cogniser is lucky enough.

Finally, justified true belief that does not constitute knowledge. Take a typical Gettier-template: I believe -justifiably, but falsely- that *a* is the only *F*; I also believe -justifiably-, that *a* is *G*. I thus go on to form the belief that the only *F* is *G*. As it happens, *b* is the only *F* and it is *G*. My belief that the only *F* is *G* is true and justified, but does not constitute knowledge.

One way to account for this possibility is as follows. Let us suppose that the following Procedure for production of beliefs is in place:

> From beliefs of the form:
>
> - *a* is the only *F*
> - *a* is *G*
>
> For any a, F and G, if conditions C obtain, form the belief
>
> - The only *F* is *G*

The conditions C are there to prevent this process from overflowing the system with innumerable useless beliefs. Only those relevant to whatever projects the cogniser has must be formed[8].

It may be that some of our belief producers follow this Procedure, which is obviously truth-conducive. It relies on the fact that if both premises are true, the conclusion will be, and on the fact that most beliefs are true. We may also assume that the two premises have been obtained through generally sound Procedures -involving testimony, maybe. There is no reason to think that all of these Procedures are working outside their ecologically-fixed context -or we may assume so for the purposes of the argument. So the Gettier belief is justified.

But KNOWLEDGE is not met: the causal grounds that explain the sufficient reliability of SUBSTITUTION are not responsible of the truth of the Gettier belief in question. As I have said, this process relies, among others, on the fact that both premises are normally true -in a sufficient number of cases-, and this is not met in our example. Thus, the Gettier belief does not constitute knowledge.

6.2 KNOWING MODAL CONTENTS

In chapter 5 (together with the rest of chapters, actually) I have sketched a strategy to naturalise modal contents. This strategy involved, first, a naturalisation of probability-involving contents and, second, an identification of an important subset of modal facts with certain quantified

6 On the other hand, there are beliefs such as *There are, or have been, HPCs*. If my account is correct, this belief is true in every world in which it is entertained by someone.

7 In my theory, beliefs that are simply put there by a neurosurgeon, or are constituted by a serendipitous brain tumour, are impossible. They lack the right history and, thus, cannot have content. Most flavours of Causal Role Semantics, e. g., would yield a different prediction.

8 Informatively stating C involves solving the frame problem, of course.

facts about probabilities. Let me stress again that this identification need not be a priori for human cognisers.

The general knowledge and justification recipes just provided may be straightforwardly applied to these contents. A judgement J that $\Diamond p$ will have the content it has in virtue of the historical properties of the belief: the function of the concatenation of producers, starting from the selected-for first link and resulting in J, together with the n^{th} order HPC that has enabled this function -and, remember: no HPC, no content. This concatenation of producers comes together with a concatenation of Procedures, and a ecologically-fixed context of application. Modal judgements made according to the concatenation of Procedures in the right contexts are justified. All other judgements are not. If the n^{th} order HPC is responsible of the existence of J and of the fact that $\Diamond p$, the judgement constitutes knowledge. As simple as that.

It is the package deal consisting of etiosemantic theory of content, branching conception of modality and ecological reliabilism that I am proposing as the right naturalistic position in modal epistemology. At the very least they provide a substantial answer to the "dubiously intelligible faculty" worry put forward (to provide yet another example) in this quote:

> Not only are we *aware* of no bodily mechanism attuned to reality's modal aspects, it is unclear how such a mechanism could work even in principle. Yablo (1993, p. 3f)

By no means am I proposing that the mechanisms I have described throughout this work are to be found in actual human brains. But I do claim I have given some detail about how a natural mechanism attuned to the modal aspects of reality could work in principle.

And this bridging of the gap between our judgements and the modal realm has not come at the expense of "some form of mind-dependence anti-realism about modality" Jenkins (n.d.)[9]. On the contrary, the conception of modality I am assuming is fully objective, and all the usual marks of realism are preserved: widespread error in our modal judgements and unknowable truths, for example, are perfectly possible.

Now that we have a summary of the view on modal epistemology that I wish to recommend which is clear enough to be assessed, even if obviously sketchy at many places, I wish to close this chapter by comparing it to another prominent approach: conceivabilism.

6.2.1 *Conceivabilism*

Many philosophers have assumed that our access to modal facts is mediated by our faculty of *conceiving*. There has been much philosophical elucidation of what it is that we mean by conceiving, with the starting point usually taken to be provided by the, altogether very natural and initially plausible, appeals to conceivability made popular by Modern philosophers. To quote from two most prominent examples, Descartes stated that God

> can bring about everything that I clearly and distinctly recognize as possible. Descartes (1641/1988, Fourth Replies, 2:154)

9 I should note that Jenkins does not defend mind-dependence about the mental.

That is, everything I (clearly and distinctly) recognize as possible is possible. Hume also famously contended that

> Whatever can be conceiv'd by a clear and distinct idea necessarily implies the possibility of existence. Hume (1740/1978, 43)

According to these two philosophers, we have an ability to conceive, the exercising of which provides insight on the modal status of propositions -*i. e.*, whether they are necessary, contingent, possible or impossible.

There is indeed an intuitive appeal to the idea that we have some faculty that, without recourse to experience, allows us to scrutinise modal properties. There are, nevertheless, two problems in this connection. One is to clarify what exactly is this ability to conceive we are talking about here. The other is to chart the scope of this purported ability: Which possibilities and necessities can we know through our conceivings? I will now briefly discuss two alternative proposals regarding the nature of our conceivings, and the scope of the access to modal facts that they allow. This will be useful for me to explain what I find lacking in conceivabilist modal epistemologies, and also to discuss in which respects the account I am developing respects conceivabilist intuitions without incurring in the same kind of problems.

Yablo's Conceivabilism

One of the best worked-out proposals about the nature of conceiving is provided in Yablo (1993). Yablo understands conceving as being a special kind of imagining, such that:

> *p* is conceivable for me if I can imagine a world that I take to verify *p*. Yablo (1993, p. 29)

That is: if I can imagine a situation of which I truly believe that *p* -Yablo (1993, p. 26). Finding something conceivable must be sharply distinguished from finding it possible-for-all-we-know. For example, suppose we get to know somehow that every mathematical statement is, if true, necessarily true and, if false, necessarily so. If we go on to consider Goldbach's Conjecture -the as yet unproven conjecture that every even number greater than 2 may be expressed as the sum of two primes-, do we deem it conceivable? According to Yablo, we do not. We simply remain undecided as regards its status from the point of view of conceivability:

> No thought experiment that I, at any rate, can perform gives me the representational appearance of the conjecture as possible or as impossible, or the slightest temptation to believe *anything* about its modal character. Yablo (1993, p. 10)

There is an appearance otherwise: we are able to imagine a situation in which newspapers and scientific journals publish the news that a certain bright mathematician has discovered a proof of the Conjecture. But this appearance is misleading; for a world to verify this imagining it is enough that it is a world in which the *false* news of the existence of a proof of the Conjecture are published in newspapers and journals.

One may retort, what about imagining a world in which the Goldbach Conjecture is proved, *tout court*? A world which verifies that imagining

needs to be a world in which the Goldbach Conjecture is true. And one cannot object to the possibility of such an imagining on the basis that there is no imagery associated to the Goldbach Conjecture being true. There is no imaginery that distinguishes an imagining of Michael Jackson climbing a flight of stairs from an imagining of a perfect double of Michael's climbing a flight of stairs. But, surely, we can imagine the latter without imagining the former scenario, or vice versa. Moreover, Yablo is not defending the mainstream option here: many conceivabilists believe that we do find both the Goldbach Conjecture and its negation, conceivable[10]. So why (or better: how) does Yablo go about denying that we find the Goldbach Conjecture conceivable?

Yablo seems to defend that conceivings justify modal judgements because they are reliable in predicting possibilities. That is, whenever we advance a positive conceivability judgement, more often than not the content we judge is possible. It is this reliability that underwrites the justification-conferring property of conceivings. So, we need to be relying on some kind of information about the proposition such that, when that information is present, we shy away from issuing a conceivability verdict.

Now, that we refrain to claim conceivability when confronted with a situation we *recognise* to be impossible is clear enough: our recognising p to be impossible does not seem to be separable from our inability to imagine a world of which p. So here we have a number of impossibilities which, as the reliability claim needs, do not go together with a positive conceivability verdict. So far, so good.

But, if conceivings are to be reliable, Yablo also needs it to be true that *non-recognised* impossibilities are such that, more often than not, we do not deem them conceivable. Otherwise, just from the fact that we don't find most *recognised* impossibilities conceivable, it would not follow that we don't find most impossibilities (*simpliciter*) conceivable: it may turn out that most impossibilities are of the non-recognised variety.

According to Yablo, luckily, it is also true that we do not deem most non-recognised impossibilities conceivable. Take any prominent examples of non-recognised impossibilities. For example, we do not recognise the Goldbach Conjecture to be impossible; nor its negation to be impossible. But we remain undecided as regards their conceivability: we are not willing to risk a judgement in which we place a creedence of, maybe, about 50%[11] -if we know, as we should, that mathematical statements are necessary if true and impossible if false. Or, if we believe that one of two women is the mother of a certain child, we are not willing to judge that each of them is conceivably his mother -having

10 According to David Chalmers, the Goldbach Conjecture is only *prima facie* conceivable. I discuss his views below.

11 Yablo believes there is no thought experiment we can perform to rationally tilt this creedence in any direction. This may well be so for Yablo, as it is for me, but how about the following case: Eva has the hobby of trying random even numbers for compliance with the Conjecture. She stumbles upon a number which her calculations show not to be the sum of any two primes (she has made a mistake,, but she is a consistently reliable mathematician, so she is perfectly justified in believing the outcome of her calculations). Believing she has produced a counterexample, she comes to believe that the Conjecture is false, whence she infers that it is possibly false.
It is natural to see that situation as one in which Eva is justified in believing the Conjecture to be possibly false.
Eva, fully rationally, may place a much higher creedence on the possibility of the negation of the Goldbach Conjecture, which is why she may well be willing to risk a judgement of conceivability.

in mind, as we should, that whoever is his mother is necessarily his mother. It is this, we may say, *Yabloan prudence* that undergirds the reliability of our conceivings.

In these examples, our reluctance to claim conceivability of the target proposition p seems to stem from our appreciation that, for all we know, we may be in one of the two[12] kinds of situations that, according to Yablo, account for modal error. One in which there is a proposition q such that

1. q
2. if q then $\Box\neg$p
3. That I find p conceivable is explained by my unawareness that 1. and/or by my unawareness that 2. Yablo (1993, p. 35, slightly edited)

So, *e. g.*, we refrain from claiming conceivability that one of two women (Martha, say) is possibly a certain child's mother because we cannot decide whether the other woman, Mary, is in fact his mother, and we recognise that our inclination to claim conceivability that p (that Martha is the mother) may stem from our unawareness that q (that Mary is), where if q then $\Box\neg$p. To avoid being in this kind of modal error, we refrain from issuing a conceivability verdict. In this way we salvage the reliability of our conceivings.

This strategy, on the other hand, seems to be vulnerable to an attack that dramatically reduces the scope of modal judgements that our conceivings can justify. Van Inwagen (1998) may be described as an attempt to show that, for any ambitious modal target proposition p, we may always identify a non-trivial epistemically-possible defeater, unawareness of which may, for all we know, underly our inclination to claim conceivability. Take, for example, our coming to believe that cows might be purple on the basis of our imagining a world of which it is true that cows are purple. A creative modal sceptic may force us to withdraw the verdict of conceivability by suggesting that the fact that we find p: *Cows are purple* conceivable stems from the fact that we are unaware of q: *There is no cow DNA that allows for purple coat pigmentation*, such that if q then $\Box\neg$p. That is, for all we know we may be victims of the kind of modal error presented above, so our Yabloan prudence kicks in and makes us withdraw our conceivability verdict.

Moreover, this strategy can be extended to the conceivability of any, however ordinary, modal proposition. Take p: *I have read Aquinas's Summa Contra Gentiles in English translation*. I am pretty sure I know p to be possible. But now consider q:

> *There is something N about my neuronal makeup such that N is a necessary property of me [maybe it is genetically necessitated by my being a son of my parents] and N makes it impossible that I read that book in particular [maybe I just cannot parse some of the very sentences that constitute the book, although I would have no problems with most other possible English sentences].*

For all I know my apparent ability to conceive of p may be due to my unawareness of the fact that q, together with the fact that if q then $\Box\neg$p.

12 The other kind of situation is one in which unawareness is substituted by denial in condition 3. below.

All in all, Yablo's approach to conceivability seems have the shortcoming of being all too sensitive to epistemically-possible defeaters to our judgements of conceivability. Otherwise put, if our conceivings are reliable only in virtue of our Yabloan prudence, then, even if a conceiver does not come up with suitable defeaters on her own, a sufficiently creative, modally-sceptic interlocutor can reduce her to silence.

Yablo has chosen this route, perhaps, because he wants to account for the reliability of our possibility judgements but does not trust any subpersonal mechanism with the task. As we have seen in a quote above, he does not believe that any bodily mechanism may be attuned to modality; instead, it may be that personal level reasoning processes that lead to what I've called *Yabloan prudence* are more plausible. But in fact, as I have just shown, such personal level processes yield too high a tax of misses -wrong rejections of candidates of possibilities. That is, their reliability is obtained at a prohibitive cost in productiveness.

It would be great if there was a mechanism which combined more wisely than us -if we are Yabloan-prudent cognisers- productivity and reliability. Luckily, there is. The kind of naturalistic modal epistemology I have been sketching in these last chapters shares with Yablo's conceivabilism the aim of providing a basis for the reliability of our modal judgements. According to my proposal, though, this reliability does not stem from Yabloan prudence in the deployment of concepts, but from the judgement producers -entirely subpersonal mechanisms such as PRED in chapter 4 or SYN in chapter 5- having stumbled upon sufficiently-reliable correlations between the cue that causes a modal judgement to be formed and the modal fact in question -which is, remember, a fact about the probability of a certain ulterior fact at certain times.

As in the example that has figured prominently in earlier chapters, one of these producers may have stumbled upon a reliable correlation between cues such as Fido's being ill-tempered and the fact that there is a high enough probability that other dogs be ill-tempered. There is no difficulty in principle about there being systems that develop a sensitivity to reliable enough cues of the states of affairs -involving probabilities and the past- that, if my suggestions in the last chapter are on the right track, constitute many metaphysical possibilities. The system subserving our own modalising capacity have developed precisely such a sensitivity.

As for the role that Yablo accords to imaginings: as I have suggested, what makes, or not, a certain modal belief or judgement justified is that it has been produced according to the correct Procedures in the correct context (see above for details). I am unsure of the role our imaginigs play in these procedures. This is, ultimately, an empirical question: do the Procedure our modalising faculty relies on involve imaginings at any crucial point, or are they just concomitant to the main process? In any event, it is plausible that at least some kinds of judgements get justified by processes involving imagery. One example Van Inwagen (*op. cit.*) gives of an everyday modal judgement we can know is: *Had we moved the table to a side, we would have had more room*. It is not implausible that the Procedure by which we get to know such things involves mental imagery of the room after moving the table; just as knowing where a city is, if it is in the center of the heel of Italy's boot, seems to involve imagery. On the other hand, it is possible that many other acts of everyday modalising -such as getting to know if I

might have been born on a different day- do not involve imaginings essentially.

Another interesting feature of the account I have been presenting is that there is no assumption that we need be aware of the grounds of the correlations between cues and modal judgements or beliefs that our modalising engines use; the workings of these mechanisms themselves may be entirely subpersonal. *Contra* Yablo (and, I think, *pro* common sense) we need to exercise no personal-level prudence for them to be reliable.

Chalmers' Conceivabilism

David Chalmers's project is more ambitious than Yablo's. He is after an *entailment* thesis between some variety of conceivability and some variety of possibility[13], as opposed to the mere reliable relation advocated by Yablo. To accomplish this, Chalmers (2002) offers an exhaustive taxonomy of varieties of conceivability, another of varieties of possibility, and a chart of the entailment relations that may be established between them.

According to Chalmers, there are three main dimensions along which conceivings may vary. With two possible values in each dimension, this makes a total of eight conceivability flavours. First, there is the positive/negative dimension:

- *Positive conceivability* Chalmers (2002, p. 150) is, to a good approximation, Yablo's notion of conceivability, as described above. Although positively conceiving involves imagining, this is supposed to be a special faculty of *modal imagination*, which transcends imagery. So, for example, according to Chalmers one may modally imagine "pairs of situations that are perceptually indistinguishable" and situations that are "unperceivable in principle".

- On the other hand, one finds a statement *negatively conceivable* Chalmers (2002, p. 149) if one cannot rule it out a priori. Negatively conceivable statements are, then, all conceptual coherent statements and only those. Negative conceivability has also been called conceptual possibility.

The second distinction is between *prima facie* and ideal conceivability. The idea here is to idealise away from the imperfections of real-world conceivers:

- One finds a statement *prima facie* conceivable Chalmers (2002, p. 147) if one finds it conceivable after some consideration. For example, after a perfunctory examination of a mathematical proof we may deem it sound, and embrace its conclusion -the conclusion is, then, prima facie conceivable, even if slightly more careful evaluation would show that the proof is not sound and the conclusion false. Also, we may imagine a very complicated sentence, full of subordinate clauses and appeals to 'the former', 'the latter', etc. such that we simply don't have the memory or attention necessary to parse it. This sentence is not prima facie conceivable.

13 In fact, his ultimate concern is providing a watertight version of the Kripkean argument against materialism defended in Kripke (1980) -see Chalmers (1996) and Chalmers (2009). Although the kind of modal epistemology I have sketched in this work has obvious implications for the conceivability argument against materialism, a satisfactory discussion thereof will have to wait.

- On the other hand, we may introduce the notion of an ideal conceiver: a cogniser who suffers of no limitations of memory, attention, time and the like[14]. She would have no problems in understanding the very complicated sentence we imagined above, and she would never ever deem sound an unsound mathematical proof. A statement is ideally conceivable if, well, the ideal conceiver finds it conceivable.

The final distinction is between primary and secondary conceivability:

- The notion of *primary conceivability* is difficult to characterise rigourously, and Chalmers has provided a very subtle and interesting elucidation in his (2002) and (2004). I suspect it may also be ultimately incoherent, but a detailed discussion of these issues would take us too far afield. In any event, for the purposes of this summary, we may identify primary conceivability with *epistemic possibility for a competent, yet sufficiently clueless cogniser*. So, *e. g.*, it is primarily conceivable that Hesperus is not Phosphorus, because there could be a competent user of HESPERUS and PHOSPHORUS who ignores that *Hesperus is Phosphorus* is true; for all that speaker knows, Hesperus may not be Phosphorus -*i. e.*, he finds *Hesperus is not Phosphorus* primarily conceivable.

  To evaluate primary conceivability we consider the world in which certain conditions obtain as being actual, and evaluate what follows -*counteractually*, as Chalmers says. For example, we imagine that it turns out (notice the indicative mood; not "it had turned out", in the subjunctive) that the watery stuff is not H_2O, and conclude then that, if so, *Water is not* H_2O is counteractually true. This is what underlies our veredict of primary conceivability to *Water is not* H_2O.

  There is a parallel notion of *primary possibility*. A statement is primary possible if it is true in any possible world *considered as actual*. For example, suppose there is indeed a possible world in which the watery stuff is not H_2O. It that is the actual world, then *water* is not H_2O: if it turns out (indicative) that the watery stuff is something else, water is that something. We may thus conclude that it is primarily possible that water not be H_2O.

- Finally, a statement is secondarily conceivable if we judge after some reflection that it might (subjunctive this time) have been possible. What kind of reflection is needed depends on the prima facie/ideal and positive/negative distinctions. Secondary conceivability is also sensible to the epistemic status of the cogniser regarding several key empirical truths. Thus, *water is* H_2O is secondarily conceivable if we judge (upon reflection) that it might have been the case that water had not been H_2O. But, if we know that water *is* H_2O, upon reflection we will not judge that such a thing might have been the case.

  And, again, there is a parallel notion of *secondary possibility*. A statement is secondarily possible if it is true in any possible world *considered as counterfactual*. There is no possible world in which

14 And which, possibly, has *ideal creativity* when it comes to finding logical/mathematical proofs. It may be argued that such a cogniser should be able to solve the Halting Problem and is, therefore, dubiously coherent. I will not press this point now.

> water is not H_2O. So it is secondarily impossible that water not be H_2O.

As I said above, Chalmers intends to use these distinctions to find an entailment thesis between some flavour of conceivability and some flavour of possibility. Chalmers's strategy could be described as an attempt at troubleshooting the Modern rationalist principle, according to which

MRP: Something is possible if and only if it is conceivable.

A host of counterexamples -of which Modern philosophers were well aware- come from our cognitive limitations: our limited memory, attention and insufficient mathematical creativity. The prima facie/ideal distinction is there to negotiate these problems: even if our limited nature makes us find conceivable some impossible scenarios, an ideal conceiver will not be subject to this problem. Also, even if some statements are too complicated or long for us to entertain them, they will pose no problems to ideal conceivers[15]. That is, we should substitute MRP with MRPIDEAL

MRPIDEAL: Something is possible if and only if it is ideally conceivable.

It is arguably this what Modern philosophers had in mind, and this what the "clear and distinct" proviso was designed to solve.

But not every counterexample to MRP comes from the lack of ideality of our conceivings. Another important source of problems for the link between conceivability and possibility is the widely accepted existence of aposteriori necessities and the (less) widely accepted existence of apriori contingencies. Take a paradigmatic example of aposteriori necessity: *Whales are mammals*. No amount of apriori reasoning, however ideal, on the part of a competent user of the concepts WHALE and MAMMAL, is going to bring out the relation between both kinds. One needs to consult the world in order to know it.

The distinction between primary and secondary conceivability is there to negotiate this second family of problems for MRP. It is true that *Whales are not mammals* is ideally conceivable, but only *primarily* so. The amount of empirical information about whalehood and mammalhood that is needed for possesion of the concept -very little, it is to be supposed- does not comment on whether whales are mammals; so there might be ideal conceivers such that, for all she knows, whales are not mammals. But, Chalmers says, nobody is defending MRP 2ARY 2ARY:

MRP 2ARY 2ARY: Something is possible if and only if it is ideally secondarily conceivable.

Where possible is to be understood as *secondarily* possible, the more or less default reading of possible. Instead, what Chalmers is defending is MRP 1ARY 1ARY:

15 A prominent objection to Chalmers's modal epistemology -Roca-Royes (forth.) calls it the *Standard Objection*- is that we cannot know what an ideal conceiver would judge about some scenario or other, not being ideal ourselves -cf. Worley (2003). As Roca-Royes points out, there are possible rejoinders available: it may be defended, for instance, that we are *locally ideal* conceivers -ideal in some restricted domain of, e. g., everyday modal claims. Even so, it would be nice to have an explanation why are we locally ideal, if we are. These kinds of explanation are sorely lacking in contemporary discussions in modal epistemology. I view my account as helping provide precisely such an explanation.

MRP 1ARY 1ARY: Something is primarily possible if and only if it is ideally primarily conceivable.

What does this amount to? Take again a case of conceiving that whales are not mammals. According to Chalmers conceiving of this is something like conceiving of a possible world in which the whaley animal around (a big, warm-blooded, aquatic animal that breathes through a blowhole) is not a mammal. Now, we have no reason to think that this world is metaphysically impossible -at least, no reason coming from the fact that whales cannot fail to be mammals. And that world being metaphysically possible is what *Whales are not mammals* being primarily possible consists in -if that *is* the actual world then, by counteractual reasoning, whales are not mammals. We have no reasons to mistrust MRP 1ARY 1ARY, so we should endorse it.

This quick summary does little justice to Chalmers's subtle and sophisticated discussion but, I think, the main thrust of his argument is captured by it. I have a general worry regarding Chalmers's approach to modal epistemology, although I can only make a very incomplete case here. Fully developing this kind of criticism will be matter for another work.

The question I would like to ask is: even if Chalmers's distinctions are ultimately successful in shielding MRP from counterexamples[16], by providing a counterexample-free alternative to the rationalist principle, is this enough to justify our allegiance to MRP 1ARY 1ARY?

In general, the mere absence of counterexamples should not, and is not, considered enough to justify allegiance to a theory. One main reason for this is the well known fact that theory is underdetermined by data. Now, what should count as data in this case is unclear. There are at least two candidates: actual truths, and well-entrenched modal beliefs. Regarding the first, it seems that the best confirmation we may expect of something being possible is its being actual[17]. Regarding the second candidate, well-entrenched modal beliefs range from truly non-negotiable everyday beliefs such as *I may have worn a different T-shirt today* to particularly compelling examples of, say, the necessity of material constitution such as *This table could not have been made of ice*.

There are many possible theories which make roughly the same predictions as MRP 1ARY 1ARY regarding these two sources of data. The theory I have been developing in this work is, arguably, one of them. But there are many others: most sensible modal epistemologies qualify. So, there must be another dimension -besides the mere absence of counterexamples- along which we may decide which, of any two equally correct theories, is the best. One obvious candidate is hinted at in Casullo (forthcoming). He there discusses several putative principles linking necessity and apriority, and he concludes, about one of them:

> [The principle that *if p is necessarily true and S's belief that p is a necessary proposition is justified then S's belief that p is a necessary proposition is justifiable a priori*] is an intuitively plausible (...) principle that enjoys no independent support but faces no decisive counterexamples. Casullo (forthcoming).

16 I have doubts about the soundness, and the usefulness, of both the prima facie/ideal and the primary/secondary distinctions, and I hope to elaborate on them in future work. For our current purposes, nevertheless, I will simply grant that they solve what they are meant to solve.

17 Of course, if the metaphysics of modality are such that the reflexivity axiom is false, this ceases to be so; but everybody believes that the modal logic correctly governing metaphysical modality is, at least, T.

The obvious tie-breaker for modal epistemologies is the possibility of offering *independent support* for one of the options. The pack consisting of a branching metaphysics of modality, an etiosemantic theory of content -specially, the part designed to deal with contents involving probabilities- and the particular kind of process reliabilism I have called ecological reliabilism provides a substantial explanation of what it is to entertain modal contents, and to be justified in believing them. This is, I think, precisely the kind of independent support Casullo finds lacking in traditional rationalist epistemologies. This is a substantial argument against the kind of conceivabilism that Chalmers defends; the reason why it is seldom made is because we don't have many proposals regarding the independent support of our modal epistemology principles. Now we have at least a sketch of one such proposal.

Of course, this comes at a price: according to the modal epistemology I defend, we are justified in believing a lot less modal contents than according to Chalmers. But this may be simply because we *are* in fact justified in believing less contents. Our desire that our modal epistemology covers more ground than it does should not cloud our judgement regarding the epistemic credentials of our theories.

7 SUMMARY AND CONCLUSIONS

In this dissertation I have aimed at sketching a naturalistic theory of mental content, *etiosemantics*, such that, first, it solves some outstanding problems with currently available accounts and, second, may be placed at the foundation of a naturalistic modal epistemology.

The first, and arguably most pressing, problem I have tackled in my dissertation (in chapter 1) is that of indeterminacy. I have shown, with the help of a formal apparatus partly borrowed from Godfrey-Smith (1996), that both producer semantics -and, among them, the version of teleosemantics with indicators that Dretske defended in Dretske (1988), cf. 1.2- and consumer semantics -and, most prominently, Millikan's biosemantics as defended in a number of books and papers since Millikan (1984) and until Millikan (2009)- are subject to some problem or other related to indeterminacy. In particular, Millikan's biosemantics can be shown to be vulnerable to the *Output problem* -see 2.2.2: the content attributions to very simple mental states that the theory warrants must involve what I have called *high* properties -properties that are closest to what the consumer needs, such as *Being food far from predators*, or some such; for high and low properties, and Input and Output problems, see 1.3.2. The appeal to natural information, in the sense of Millikan (2004), can lower the properties that contents must involve, but not down to the point in which the content attribution follows our intuitions. Of course, our intuitions about the right content attribution to the states of very simple mechanisms in the brain of idealised animals are not decisive evidence for or against a theory, but they *are* partial evidence in favour of the theory that conforms with them. And how does etiosemantic go about solving the Indeterminacy Problem?

The key is taking seriously the very plausible idea that the entity a certain mental state is about should lie somewhere causally upstream from the representation -this I have called the COMPRESSED EXPLANATION principle. Both causal theories of content and teleosemantics abide by this principle, but the position they accord to the entity represented in the causal history leading to the representation is not enough to fix a univocal content. Such a history -in the case of a selected representation, which is what we are dealing with in chapters 1 and 2- involves the frequent coinstantiation of a mix of high and low properties such that there is a causal structure that ensures that the mix keeps recurring. The proposal is that the entity the representation is about is *everywhere* in which the causal structure in question is making the mix of properties occur. I take some care to describe exactly which is the nature of this kind of entities, which I call, with a nod to Boyd (1988), *Homeostatic Property Clusters*, or HPCs. For a more detailed discussion of HPCs in my sense see sections 1.4.5, 1.4.6 and 2.5.2. HPCs are, with some idealisation, the kind of entities we want very simple contentful states to be about, entities which lie at the origin of both detectable properties and useful properties and which, therefore, explain how detectors bring usefulness. Once we look at matters closely it is very plausible that these entities are what the most primitive contents are about.

We cannot do many things with selected representations. At any rate, most of our own representations, which is what we are most interested in, are not selected but *ephemeral:* they exist for the first and last time, during the lifetime of a not particularly longevous individual. I tackle ephemeral mental states from chapter 3 on. I recover Millikan's idea that an ephemeral state has content only if it has been produced by chain of mechanisms (possibly only one) that bottoms out in a selected mechanism. Now, Millikan develops this idea by introducing *adapted* and *derived* functions: functions of the ephemeral states or mechanisms that may help fix their content. I discuss this idea in 3.2, and conclude that the normative dimension of the behaviour of ephemeral states and mechanisms has not been conclusively established by Millikan. Her theory of derived functions seems to hesitate between the mere introduction of terminology -according to this strand, the derived functions of an ephemeral mechanism would not be anything over and above the relational function of its selected producer- and the postulation of additional normativity at the level of ephemeral mechanisms. In my discussion I suggest that one cannot get additional normativity out of terminological manipulation and, thus, that the appeal to the normativity of ephemeral mechanisms has not been discharged in a naturalistically acceptable way.

Luckily, we do not need a robust notion of the function of ephemeral states and mechanisms to develop an account of their content. I try to develop an alternative in 3.3 and 3.8. The main idea is that selected producers fix the causal profile of their products, and these products enter in relations with HPCs that mimic, in the short term, the relation that selected representations establish with the HPCs they are about. To avoid indeterminacy in the content of ephemeral mechanisms, though, we need the fact that the producer has been selected to be explained by what I call a *higher order HPC*: one that explains the presence, in the right positions in the history of this mechanism, of the lower order HPCs the ephemeral representations are about. In a nutshell, if we are to credit ephemeral mental states with content, the graph that links these states with its selected producer must be mirrored in the world, with higher order and lower order HPCs linked in the same way.

This strategy for the attribution of content to ephemeral states is applicable to very different contents and mental structures; in particular, in chapter 3 I discuss the idea in connection with two examples. First, I apply it to the case of contents which record causal connections among properties -I find this case interesting for two reasons: such contents may be precursors of thoughts in which a property is predicated of an individual, and there is a real-life mechanism whose products may be attributed with just this kinds of contents: long term potentiation (see 3.4). Second, to the case of contents involving individuals -although individual-involving contents need not necessarily be ephemeral, and examples to the contrary are provided in 3.7, most are. I also discuss a real life example, whether what ethologists call *individual recognition* is present in American lobsters, that depends, I claim, on the existence of individual-involving ephemeral states in lobsters. A fuller discussion of the consequences of this for the debate about individual recognition in ethology can be found in Appendix C.

In the first three chapters, then, I have explained how to attribute content to some selected and some ephemeral representations. So far, though, these representations show an almost complete lack of structure:

in the kind of mental states with the content *There is an F around* we have discussed, there are no components with the content *F* or *around*. Now, we need to account for representations with subpropositional contents (say, concepts; or mechanisms that effect predications in thought) if we are to account for the representations of sophisticated content-crunching engines such as human beings. This is the task I take up in chapter 4.

What I am after in this chapter is a characterisation of a couple of mechanisms that should be part of a small (maybe the smallest) cognitive system capable of productivity in thought: A mechanism, CONC, that is able to output new concepts and a mechanism, PRED, that is able to conjoin concepts in a predicative thought. The idea is that the lessons learned from the implementation of these mechanisms could be used in characterising other, more complicated mechanisms in more complicated cognitive systems, although I have not taken up this further task here.

Characterising CONC and PRED is complicated, among other things, because the most simple contentful states *have* to be propositional: one of the reasons is that we need to make room for the fact that concepts, say, FLY, may be tokened in the absence of their reference in circumstances in which doing so is normatively impeccable; e. g., while thinking *I saw a fly yesterday*, or when thinking about halteres and metathoraxes leads me to think about flies. Instead, it is not possible to token a proto-belief with the content *There is a fly around* which is normatively impeccable if there is no fly around -because it will be false. See 4.1 for further discussion. Hence the need to start from propositional contents, and hence the problem of extracting subpropositional contents out of those. Before advancing my own positive proposal, I review Millikan's attempt at accounting for compositionality, and raise some problems with it. Millikan's idea is that thoughts form a sign system, in which certain transformations of thoughts correspond to certain transformations of states of affairs. I suggest that the existence of this mapping function cannot be previous to the existence of a fully compositional language of thought, as it would have to be if it is to play the role Millikan accords to it in her theory. On the other hand, the appeal to mapping functions does not seem to be perfectly naturalistic, in the sense that there are no naturalistic grounds to choose among different candidate mapping functions which overlap in the domain causally needed for the selection of a sign system -see 4.5.

My proposal, as I said above, takes seriously the need for a compositional semantics in thought. The way to make this compatible with the fact that the most basic representations are propositional is what I call the *interlocking determination* of a predication-producing mechanism, PRED, a concept-creating mechanism, CONC, and an initial pool of thoughts produced by these mechanisms -see 4.4.2. The fact that thoughts in the pool have the content they have helps fix the way in which PRED creates thoughts, and the way in which CONC creates concepts. But, *pace* Millikan, we do not need all the correspondences to be fixated beforehand, and they might well be a partly haphazard consequence of the initial pool of thoughts and the causal profile of PRED and CONC. And, as Fodor (2008) would want, this is compatible with a punctuated mind -a mind capable of only thinking one thought. In 4.4.2, in fact, I show how the theory of concepts that emerges from the previous discussion deals (satisfactorily, I think) with many of the objections against other prominent theories of concepts.

Chapter 4 ends the part of the dissertation dedicated to mental content. In briefest summary: the content of selected atomic representations depends on an HPC occupying key positions causally upstream (chapters 1 and 2); the content of ephemeral atomic representations depends on a hierarchy of HPCs covering the behaviour of a hierarchy of producers and products (chapter 3); the content of ephemeral representations exhibiting compositionality depends on these two features, together with the interlocking determination of thought-producing and concept-producing mechanisms (chapter 4).

In chapter 5 I apply the ideas developed in part I to the case of modal contents, and finally, in chapter 6 I explain how the etiosemantic account of content may be used as the foundation for a naturalistic epistemology and, in particular, how modal contents may be justifiedly believed or known. Chapter 5 defends a metaphysics of modality (the *branching conception* of modality) according to which a sizeable portion of the modal space (maybe all of it, but I wish to remain uncommitted) is constituted by states of affairs such as *at some time it was an open possibility that p*, where something is an open possibility if there is a probability greater than zero of it happening -see 5.6. In the chapter I have taken for granted that etiosemantics is able to naturalise contents involving past times, and the existential quantifier. Now, while the first bit is, I dare think, uncontroversial, I have said nothing, or very little, about how the naturalisation of quantifiers may go. I hope to tackle this issue in future work, which so far remains a substantial loose end in the overall picture. Anyway, granting that the naturalisation in question may be done, there remains another important family of contents about which I have said nothing up until this chapter, and which are, arguably, about the distinctively *modal* ingredient in the branching conception of modality: contents involving probabilities.

The strategy for the naturalisation of probability-involving contents in the chapter is showing how the selection of a mechanism could make essential use of the probabilistic relation among a couple of properties. The mechanism in question takes profit of this probabilistic relation to issue a cost-effective response to one of them that still maximises the prospects of success in dealing with the other -for details, see 5.2 and the following sections. Mechanisms such as these I have called *synergic associations*. It is synergic associations themselves that bear contents involving probabilities. Selected synergic associations are in the family of selected proto-beliefs discussed in 3.1; to account for ephemeral contentful states involving probabilities I have characterised a particular class of synergic association producers -5.4. This is not to say that all modal contents have to be mediated by synergic associations. I don't think they do. What I think is, first, that these kinds of associations are among the simplest bearers of contents involving probabilities. And, second and most important, they provide a solid rejoinder to an often voiced objection against efforts in the naturalisation of modal semantics and epistemology, according to which it is simply unintelligible how mechanisms might be (in Yablo's turn of phrase) attuned to the modal aspects of reality. Well, synergic associations is, at least, one way in which they might. The argument from the impossibility of a mechanism sensitive to modality may be countered simply by showing a way in which such mechanism may exist. Such a mechanism is as paradoxical as a mechanism attuned to future aspects of reality, which are also causally inert now -but nobody thinks it impossible or unintelligible to

develop a naturalistic semantics of future events. The trick, as always, is have the mechanism "learn from experience": if an event of type A is consistently followed by an event of type B, a mechanism may start exploiting this relation to prepare itself for events of kind B: sensitivity to the future. If events of type B follow probabilistically from events of type A, a synergic association may exploit this relation to achive a cost-effective preparation for events of kind B: sensitivity to probabilities.

The last chapter of my dissertation is dedicated to sketching a naturalistic modal epistemology. In fact, the contention is that modal epistemology is not all that different from the epistemology of what is actually the case. This falls out from the fact that modal contents are not all that different from contents involving what is actually the case, together with the particular brand of process reliabilism -the theory according to which, roughly, a belief is justified if it was produced by a reliable process- I defend in this chapter. I take up the famous complaint that process reliabilism is not an assessable theory until there is a principled way to choose the right principle for each belief -the *Generality Problem*. Etiosemantics has a natural proposal on offer: every contentful state, however ephemeral, is linked to a selected mechanism via a chain of mechanisms (possibly only one). Each of these mechanisms imparts a Procedure upon its product -see 3.6. This chain of Procedures, moreover, is defined within what I call an *ecologically-fixed context*. Beliefs are justified if they are produced in their ecologically-fixed context -see 6.1.1. This provides a principled way to choose the process in question: the chain of Procedures; and a context in which it is reliable: the ecologically-fixed context.

This sketch of a theory, which I have called *ecological reliabilism*, is straightforwardly applicable to modal contents, which are also the product of a chain of Procedures, in an ecologically-fixed context. This is going to explain that, most of the times, we are justified in believing (and know) the everyday modal claims that strike us as true -see 6.2. That is, an explanation of what Van Inwagen (1998) calls our everyday powers of modalising is, I dare say, well within the reach of the theory formed by etiosemantics plus the branching conception of modality plus ecological reliabilism. This meets, at least to a certain degree of detail, the main goal of the dissertation. In the final section I also discuss briefly a prominent neo-rationalist approach to modal epistemology, conceivabilism, and try to show why my proposal is to be preferred.

Part III

APPENDICES

A

THE DERIVATION OF THE INDETERMINACY PROBLEM

THE CONVERSION AMONG FITNESS VALUES

Let us take up again the case discussed in 1.3.1. We have a mental mechanism M that indicates, among other properties, *Being a fly* (F) and *Being a black speck* (G). We have the values of the Fitness Matrix as seen from the perspective of the indication of instantiations of F (i. e., FM_F) and we wish to know the values of the Fitness Matrix as seen from the perspective of the indication of instantiations of G (i. e., FM_G).

Let us calculate w_{11}^G, assuming that all flies are black specks ($P(G|F) = 1$). A hit when indicating black specks will have the fitness value of a hit when indicating flies if the black speck is a fly, and the fitness value of a false alarm when indicating flies if it is not. That is, w_{11}^G is a linear combination of w_{11}^F and w_{21}^F. We simply need to average by the amount of hits that are of one type or the other:

$$w_{11}^G = \frac{P(on \wedge F \wedge G)}{P(on \wedge G)} w_{11}^F + \frac{P(on \wedge \neg F \wedge G)}{P(on \wedge G)} w_{21}^F$$

Now, according to the definition of conditional probability, $P(a|b)\,P(b) = P(a \wedge b)$. So $P(on \wedge F \wedge G) = P(on|F \wedge G)\,P(F \wedge G)$, which, if $P(G|F) = 1$, is equal to $P(on|F)\,P(F \wedge G)$. While $P(on \wedge G) = P(on|G)\,P(G)$.

On the other hand, $P(on \wedge \neg F \wedge G) = P(on|\neg F \wedge G)\,P(\neg F \wedge G)$, which is obviously equivalent to $P(on|\neg F \wedge G)\,P(\neg F \wedge G \wedge G)$. Putting it all together,

$$w_{11}^G = \frac{P(on|F)\,P(F \wedge G)}{P(on|G)\,P(G)} w_{11}^F + \frac{P(on|\neg F \wedge G)\,P(\neg F \wedge G \wedge G)}{P(on|G)\,P(G)} w_{21}^F$$

Again, using the definition of conditional probability:

$$w_{11}^G = \frac{P(on|F)}{P(on|G)} P(F|G)\, w_{11}^F + \frac{P(on|\neg F \wedge G)}{P(on|G)} P(\neg F \wedge G|G)\, w_{21}^F$$

Which, after reorganising, gives

$$w_{11}^G = \frac{1}{P(on|G)} \left[P(F|G)\,P(on|F)\, w_{11}^F + P(G \wedge \neg F|G)\,P(on|G \wedge \neg F)\, w_{21}^F \right]$$

To calculate w_{12}^G we proceed in an analogous manner.

FITNESS CONTRIBUTIONS

M's Fitness Contribution as seen from the perspective of the indication of black specks (FC^G) is

$$\begin{aligned} FC^G &= P(G)\cdot\left[P(on|G)\,w_{11}^G + P(off|G)\,w_{12}^G\right] + \\ &+ P(\neg G)\cdot\left[P(on|\neg G)\,w_{21}^G + P(off|\neg G)\,w_{22}^G\right] \end{aligned}$$

First, we substitute w_{ij}^G by their equivalences in terms of w_{kl}^F - simplifying $\frac{P(on|G)}{P(on|G)}$ and $\frac{P(off|G)}{P(off|G)}$ along the way.

$$\begin{aligned} FC^G &= P(G)\cdot \\ &\quad \left[P(F|G)\,P(on|F)\,w_{11}^F + P(G\wedge\neg F|G)\,P(on|G\wedge\neg F)\,w_{21}^F + \right. \\ &\quad \left. P(F|G)\,P(off|F)\,w_{12}^F + P(G\wedge\neg F|G)\,P(off|G\wedge\neg F)\,w_{22}^G\right] + \\ &+ P(\neg G)\cdot\left[P(on|\neg G)\,w_{21}^F + P(off|\neg G)\,w_{22}^F\right] \end{aligned}$$

Now we resolve brackets and reorganise.

$$\begin{aligned} FC^G = {} & P(G)\,P(F|G)\,P(on|F)\,w_{11}^F + P(G)\,P(F|G)\,P(off|F)\,w_{12}^F + \\ & P(\neg G)\,P(off|\neg G)\,w_{22}^F + P(G)\,P(G\wedge\neg F|G)\,P(off|G\wedge\neg F)\,w_{22}^G + \\ & P(\neg G)\,P(on|\neg G)\,w_{21}^F + P(G)\,P(G\wedge\neg F|G)\,P(on|G\wedge\neg F)\,w_{21}^F \end{aligned} \tag{A.1}$$

We should now take into account that

$$P(A)\,P(C|A) + P(B)\,P(C|B) = P(A\wedge B)\,P(C|A\wedge B)$$

and, in particular, given that $\neg F \leftrightarrow \neg G \vee (G\wedge\neg F)$, we have that

$$P(G\wedge\neg F)\,P(on|G\wedge\neg F) + P(\neg G)\,P(on|\neg G) = P(\neg F)\,P(on|\neg F) \tag{A.2}$$

and

$$P(G\wedge\neg F)\,P(off|G\wedge\neg F) + P(\neg G)\,P(off|\neg G) = P(\neg F)\,P(off|\neg F) \tag{A.3}$$

We also know that

$$P(G)\,P(F|G) = P(F\wedge G) = P(F) \tag{A.4}$$

Substituting the results A.2, A.3 and A.4 in equation A.1, we get:

$$\begin{aligned} FC^G &= P(F)\,P(on|F)\,w_{11}^F + P(F)\,P(off|F)\,w_{12}^F + \\ &\quad P(\neg F)\,P(off|\neg F)\,w_{22}^F + P(\neg F)\,P(on|\neg F)\,w_{21}^F \\ &= P(F)\left[P(on|F)\,w_{11}^F + P(off|F)\,w_{12}^F\right] + \\ &\quad P(\neg F)\left[P(on|\neg F)\,w_{21}^F + P(off|\neg F)\,w_{22}^F\right] \end{aligned}$$

That is, $FC^G = FC^F$, as we wanted to show.

B

CONDITIONS FOR SUCCESSFUL RECRUITMENT

Suppose M is a G-mechanism (i.e., a mechanism such that M's being *on* has the content *There is a G around*) which has some property E as its input: the instantiation of E around M's possessor causes M to go *on*, and nothing else does. I will characterise when is it useful to recruit another property F for M's input.

Before recruiting F, M fires when and only when E is instantiated. Thus, its Indication Profile from the perspective of the indication of Gs is equivalent to:

$$IP^G = \begin{bmatrix} P(E|G) & P(\neg E|G) \\ P(E|\neg G) & P(\neg E|\neg G) \end{bmatrix}$$

and, thus,

$$FC^G = P(G) \cdot [P(E|G)\, w_{11} + P(\neg E|G)\, w_{12}] + P(\neg G) \cdot [P(E|\neg G)\, w_{21} + P(\neg E|\neg G)\, w_{22}]$$

DISJUNCTIVE RECRUITMENT OF A PROPERTY

One way of recruiting the real kind F for the input of the M is making M fire whenever either E or F are around. We call this *disjunctive recruitment*.

In this situation, the indication profile of M will change to:

$$IP^{\vee} = \begin{bmatrix} P(E \vee F|G) & P(\neg (E \vee F)|G) \\ P(E \vee F|\neg G) & P(\neg (E \vee F)|\neg G) \end{bmatrix}$$

The fitness contribution of M will change accordingly:

$$\begin{aligned} FC^{\vee} \quad = \quad & P(G) \cdot [P(E \vee F|G)\, w_{11} + P(\neg (E \vee F)|G)\, w_{12}] + \\ & P(\neg G) \cdot [P(E \vee F|\neg G)\, w_{21} + P(\neg (E \vee F)|\neg G)\, w_{22}] \end{aligned}$$

It is rewarding to recruit F disjunctively for M's input if and only if said recruiting amounts to an increase in fitness contribution. That is, iff $FC^{\vee} - FC^G > 0$. Take into account that

- $P(E \vee F|G) - P(E|G) = P(F \wedge \neg E|G)$. The probability of either E or F being instantiated, conditional on G, minus the probability of E being instantiated conditional on G, is the probability of F being instantiated in the absence of E, conditional on G.
- $P(\neg E|G) - P(\neg (E \vee F)|G) = P(F \wedge \neg E|G)$. The probability of E not being instantiated, conditional on G, minus the probability of neither E nor F being instantiated conditional on G, is the probability of F being instantiated in the absence of E, conditional on G.
- $P(E \vee F|\neg G) - P(E|\neg G) = P(F \wedge \neg E|\neg G)$
- $P(\neg E|\neg G) - P(\neg (E \vee F)|\neg G) = P(F \wedge \neg E|\neg G)$

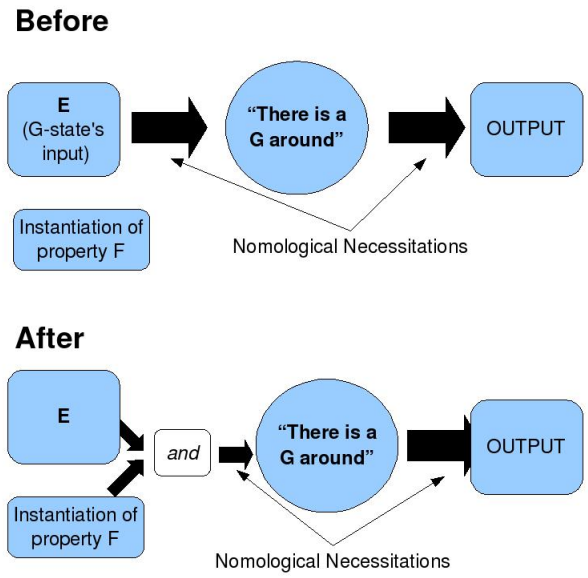

Figure 17: Conjunctive Property Recruitment

And, so,

$$\begin{aligned} FC^{\vee} - FC^{G} &= P(G) \cdot P(F \wedge \neg E|G)(w_{11} - w_{12}) - \\ &\quad P(\neg G) \cdot P(F \wedge \neg E|\neg G)(w_{22} - w_{21}) > 0 \end{aligned}$$

We may call $P(F \wedge \neg E|G)$ F's *positive excess* over E. This positive excess records the probability of M now firing in the presence of G in cases in which it would have not fired with E as its only input property. The positive excess yields a higher probability of cashing in the benefit of hits (w_{11}) and a decreased probability of incurring in the cost of misses (w_{12}). Contrariwise, $P(F \wedge \neg E|\neg G)$ is F's *negative excess* over E. M, with its new input, will overfire with a probability recorded by the negative excess. This implies a lower rate of correct rejections (w_{22}) and a higher rate of false alarms (w_{21}).

CONJUNCTIVE RECRUITMENT OF A PROPERTY

Another way to recruit F is to make M fire when and only when F and E are instantiated together.

That is, moving to the following indication profile:

$$IP^{\wedge} = \begin{bmatrix} P(E \wedge F|G) & P(\neg(E \wedge F)|G) \\ P(E \wedge F|\neg G) & P(\neg(E \wedge F)|\neg G) \end{bmatrix}$$

By a reasoning parallel to the previous case, such a *conjunctive recruiting* is rewarding iff

$$\begin{aligned} FC^{\wedge} - FC &= P(\neg G) \cdot P(E \wedge \neg F|\neg G)(w_{22} - w_{21}) - \\ &\quad P(G) \cdot P(E \wedge \neg F|G)(w_{11} - w_{12}) > 0 \end{aligned}$$

Where $P(E \wedge \neg F|\neg G)$ may be called F's *positive defect* over E and $P(E \wedge \neg F|G)$ its *negative defect*.

Disjunctive recruiting will be preferred in situations in which false alarms are not very costly and hits are very rewarding (the import of "very" and "not very" here is made precise by $FC^{\vee} - FC$). An example may be a situation in which nutritional resources are scarce: it may pay to overfire more if it means missing less prey. Conjunctive recruiting will be preferred in situations in which false positives are very costly. Other, more complicated patterns of integration of sensitivities to E and F may be preferred in the presence of more symmetrical fitness matrixes.

RECRUITING F WHEN ALL GS ARE FS

Consider the case in which the property F which is to be recruited for the input of the G-mechanism is (in situ) necessary for G to be instantiated around the possessor of the mental state. That is, $P(F|G) \approx 1$. In this situation, $P(F \wedge \neg E|G) \approx 1 - P(E|G)$ and $P(E \wedge \neg F|G) \approx 0$. Nothing can be said in general about the positive defect and negative excess of F over E. The conditions for disjunctive or conjunctive recruiting to be rewarding are:

$$\begin{aligned} FC^{\vee} - FC &= P(G) \cdot (1 - P(E|G)) \cdot (w_{11} - w_{12}) - \\ &\quad P(\neg G) \cdot P(F \wedge \neg E|\neg G)\,(w_{22} - w_{21}) > 0 \end{aligned}$$

$$FC^{\wedge} - FC = P(\neg G) \cdot P(E \wedge \neg F|\neg G)\,(w_{22} - w_{21}) > 0$$

Conjunctive recruiting will always be rewarding, on condition that $P(E \wedge \neg F|\neg G) > 0$, which amounts, in this case, to saying that F must not be always and everywhere instantiated, and E is not sufficient for G. This is just to be expected from the fact that F is necessary for G.

The condition for disjunctive recruiting to be rewarding, on the other hand, is more difficult to meet, although it may surely obtain in some cases.

RECRUITING F WHEN ALL FS ARE GS

Take now the case that $P(G|F) \approx 1$, then

$$FC^{\vee} - FC = P(G) \cdot P(F \wedge \neg E|G)\,(w_{11} - w_{12}) > 0$$

$$\begin{aligned} FC^{\wedge} - FC &= P(\neg G) \cdot (1 - P(\neg F \wedge \neg E|\neg G))\,(w_{22} - w_{21}) - \\ &\quad P(G) \cdot P(E \wedge \neg F|G)\,(w_{11} - w_{12}) > 0 \end{aligned}$$

Here it is disjunctive recruiting that will always be rewarding, provided that E is not necessary for G, and F is somewhere and sometime instantiated. Conjunctive recruiting has a more complicated condition.

THE CONTENT OF M AFTER RECRUITMENT

As we have already seen in previous chapters, the content of a state has to do with the causal underpinnings of the conditional probabilities that are instrumental in the explanation of the actual existence of the

state. Will the content of a G-mechanism change after recruiting of some property F for its input?

It is to be expected that, in most cases, if a property F increases the Fitness Contribution of some state, it is because instantiations of F are (were all along) part of the Cluster of the HPC that constitutes property G. In those cases, M does not change content after recruitment -I'm not saying, but should always be taken to imply: only if the recruitment is subsequently selected for.

It is not impossible, though, that F taps into another source of fitness -*i. e.*, indicating F makes M fitness-conducive, but for reasons which are independent from G. We could imagine that F makes M track more, and different prey. In this case, the final content of M may be a disjunction of G and some other property that grounds the improvement in fitness contribution supplied by F.

DISJUNCTIVELY RECRUITING AN F-MECHANISM

Suppose we have an F-mechanism and a G-mechanism in a case in which $P(F|G) = 1$. Let us calculate the Fitness Contribution of both states before, and after, disjunctively recruiting the F-mechanism's firings for the input of the G-mechanism:

BEFORE.

- The input to the F-mechanism is some property E. Therefore, its Indication Profile is $IP_F = \begin{bmatrix} P(E|F) & P(\neg E|F) \\ P(E|\neg F) & P(\neg E|\neg F) \end{bmatrix}$
- The input to the G-mechanism is some property H. Therefore, $IP_G = \begin{bmatrix} P(H|G) & P(\neg H|G) \\ P(H|\neg G) & P(\neg H|\neg G) \end{bmatrix}$
- The fitness matrix for the F-mechanism is $FM_F = \begin{bmatrix} w^F_{11} & w^F_{12} \\ w^F_{21} & w^F_{22} \end{bmatrix}$
- The fitness matrix for the G-mechanism is $FM_G = \begin{bmatrix} w^G_{11} & w^G_{12} \\ w^G_{21} & w^G_{22} \end{bmatrix}$
- Fitness Contributions are, then:

$$\begin{aligned} FC^F &= P(F) \cdot \left[P(E|F)\, w^F_{11} + P(\neg E|F)\, w^F_{12} \right] + \\ &\quad P(\neg F) \cdot \left[P(E|\neg F)\, w^F_{21} + P(\neg E|\neg F)\, w^F_{22} \right] \end{aligned}$$

$$\begin{aligned} FC^G &= P(G) \cdot \left[P(H|G)\, w^G_{11} + P(\neg H|G)\, w^G_{12} \right] + \\ &\quad P(\neg G) \cdot \left[P(H|\neg G)\, w^G_{21} + P(\neg H|\neg G)\, w^G_{22} \right] \end{aligned}$$

- The total Fitness Contribution of both mechanisms to the agent is just $FC^{F+G} = FC^F + FC^G$

AFTER. After recruitment, the contribution of the F-mechanism remains the same, but the G-mechanism changes:

- $IP^{G*} = \begin{bmatrix} P(H \vee E|G) & P(\neg(H \vee E)|G) \\ P(H \vee E|\neg G) & P(\neg(H \vee E)|\neg G) \end{bmatrix}$
-

$$\begin{aligned} FC^{G*} &= FC_G + P(G) \cdot P(E \wedge \neg H|G) \left(w_{11}^G - w_{12}^G\right) - \\ &\quad P(\neg G) \cdot P(E \wedge \neg H|\neg G) \left(w_{22}^G - w_{21}^G\right) \end{aligned}$$

Recruitment of the F-mechanism will be rewarding iff $FC_G^* > FC_G$. In that case we say that $P(G|F)$ is *sufficiently* higher* than $P(G)$for the agent. Thus

- $FC^{F+G*} = FC^F + FC^G + P(G) \cdot P(E \wedge \neg H|G) \left(w_{11}^G - w_{12}^G\right) - P(\neg G) \cdot P(E \wedge \neg H|\neg G) \left(w_{22}^G - w_{21}^G\right)$. The third addend is the net contribution of the recruitment, and is positive for normal fitness matrices -in which $w_{11}^G > w_{12}^G$ and $w_{22}^G > w_{21}^G$.

Another way to look at the recruitment is as follows: the G-mechanism remains unchanged, and the fitness matrix of the F-mechanism changes to cash in the extra benefit obtained by the G-mechanism. The final result of FC^{F+G} is, of course, the same.

C

INDIVIDUAL RECOGNITION WITHOUT SPECIFICITY

There is an ongoing debate in ethology concerning the true nature of Individual Recognition [IR, from now on]: in what circumstances is some animal recognising a concrete individual as such individual, as opposed to, say, as a family member (this would be kin recognition) or as a member of some kind or other?

It has been suggested (*e. g.*, in Tibbetts and Dale (2008) and Tibbetts et al. (2008)) that, if we are to credit a receiver with the ability of recognising individual signallers, all three of the following features must be individual-specific: the cue produced by the signaller [cue]; the template the receiver uses to match the cue [template]; the behavioural response to the cue [response]. On the other hand, others (e. g. Steiger and Müller (2008)) defend that only cue and template should be individual specific.

I intend to show that, appearances notwithstanding, *neither* cue nor response must be individual-specific for IR to exist. On the one hand, true specificity in any of these two features is extremely costly, if not a downright myth, impossible to come by. On the other hand, less than perfectly specific cues may show sufficient correlation with concrete individuals to be useful for the purposes the receiver puts them to. Instead of specificity, a reasonable notion of IR must make appeal to the role a concrete individual plays in the explanation of the existence of a certain representation in the receiver's cognitive economy -along the lines discussed in chapter 3.

Cue Specificity.

Most discussions of IR in the literature in evolutionary biology (*e. g.*, Müller et al. (2003), Tibbetts and Dale (2008), Steiger and Müller (2008), Tibbetts et al. (2008), an exception is Sherman et al. (1997)) assume that, if we are to credit some agent with recognition of individuals, the cues the former use must be *individual specific*. The first thing to point out is that it is hard to see what precisely individual-specificity must amount to. Honest-to-God individual specificity will not do:

TRUE SPECIFICITY: Some cue, C, is specific to some individual, I, iff $P(I|C) = 1$.

and

IR - SPECIFICITY: A receiver is able to perform IR only if it uses individual-specific cues in the sense of TRUE SPECIFICITY.

Individual-specific cues in this sense do not exist, or at least are extremely hard to come by[1]. A common experimental setup for testing

1 There is a sense in which every cue is individual-specific, and more than that: the fine physical structure of cues is sufficient, most probably, to distinguish every cue -even leaving aside its causal-historical properties- from every other cue and, for example, the call of baby seals and recordings of the very same token calls must undoubtedly differ in subtle ways -their frequency spectrum, say. Biologists participating in the IR debate take for granted a common-sense grouping of cues according to which the call of baby seals and a recording of those calls count, naturally enough, as the same cue. I will go along with them in assuming the existence of such a common-sense grouping. In the main text I'm saying that there are no individual-specific cues in this sense.

IR (as described in Tibbetts and Dale (2008, p. 529)) relies precisely on IR - SPECIFICITY being *false*: presenting the receiver with a putative cue and observing its response would be impossible if the presence of the cue made necessary the presence of the cued individual. But, of course, a researcher may, *e. g.*, record the call of a baby seal and play it back to an adult individual to observe its responses; this already shows that the probability of the presence of the cue conditional on the absence of the individual is, in many situations in which we attribute IR to a receiver, not zero -and, therefore, shows that $P(I|C) < 1$ in cases of IR.

Counterexamples to IR - SPECIFICITY are possible also without the intervention of a researcher. Human facial traits are not truly individual-specific according to the definition above: it is just extremely unlikely, but by no means nomologically impossible, that two different persons look indistinguishably alike to us. Some very young homozigotic twin siblings are actual examples of this situation. But, even if we may sometimes mistake one twin sibling for the other, it is indisputable that human facial traits are a good-enough cue for purposes of IR.

Response Specificity.

On the other hand, Tibbetts et al. (2008), identify response-specificity with cases in which a receiver "treats [an] individual differently from others" (*op. cit.*) This may be read as meaning either of two things, none of which are very promising:

ALL: There is response-specificity to an individual I only if the response to I is different from the response to *all* other individuals.

SOME: There is response-specificity to an individual I only if the response to I is different from the response to *some* other individuals.

There are several relatively uncontroversial instances of IR that appear to provide counterexamples to the response-specificity requirement, if understood as in ALL. So, *e. g.*, the discussion in Karavanich and Atema (1998) of dominance in American lobsters hypothesises that lobsters are capable of IR, even if they only show two different responses in their encounters with other lobsters: flee or fight. These two responses are highly non-specific: avoidance is the right behaviour to display against *any* individual placed higher-up in the dominance hierarchy; fighting is the right behaviour to display against any individual in a lower position. It is by no means true that a lobster L_j's response to lobster L_i is different from its response to all other lobsters.

The reading in SOME does account for cases such as this: L_j's response to L_i is different from its response to some other lobsters: those that occupy a lower position than L_j in the hierarchy, if L_i is higher than L_j; or those that occupy a higher position, if L_i is lower. But, then, it is unlikely that anybody disagree that IR involves response-specificity in this sense. After all, denying this would be equivalent to defending that IR may be present even if the response is the same for all individuals. This is a position nobody holds. In summary, in the one hand ALL gives a substantive understanding of response-specificity; but there are cases of IR without response-specificity in this sense. On the other hand, SOME accounts for these cases, but it is not a substantive notion of response-specificity -or, at least, nobody wishes to deny that there is response-specificity in this weaker sense.

A Better Analysis of IR.

Likely, high cue- and response-specificity are well correlated with the presence of IR but, as I have just shown, specificity is not necessary for IR in either case. Luckily, there is a better way to provide necessary and sufficient conditions for the presence of IR:

IR - CONTENT: A receiver is able to perform IR if and only if it is capable of having mental states with contents of the kind $F(a)$, where *a* is an individual.

This is, quite simply, the common-sense way to analise IR. The reason ethologists shy away from something like IR - CONTENT is, probably, because they do not believe the notion of content to be sharp enough to play a role in our scientific theories about animal behaviour. The good news is that, while we certainly do not have, and maybe there are not, sets of necessary and sufficient conditions for a mental state to have a content of the kind $F(a)$, I have provided -in chapter 3- an interesting set of sufficient conditions for the (pretty central) case of ephemeral states and the content *a is around*:

A IS AROUND - EPHEMERAL: An agent A has a mechanism, M_i, whose positives have the content a_i *is around*, if cue C_i has caused N to create it and all of ROUTINE, FITNESS CONDUCIVENESS and HIGHER ORDER HPC are in place -see 3.8 for details.

A IS AROUND - EPHEMERAL makes the question whether ephemeral states have the content *A is around* sufficiently well-formulated for it to figure in a description of the state of affairs that constitutes IR. How does this proposal relate to the traditional idea of having IR depend on cue- or response-specificity? It can be shown that, in general, the specificity of cues to individuals must be high enough to make the practice of IR self-sustaining. That is, it must be high enough to make N fitness-conducive in the long run: one of the conditions for attributing content to N's products is that, in a sufficient number of such products M_i,

FC1: For a number $j \geqslant 1$ of properties G_j, M_i has indicated the instantiation of G_j around its possessor.

Between these properties G_j that M_i indicates we have those that constitute, together with a specialised homeostatic mechanism, the individual M_i is about. So, suppose that we introduce a measure of *state-specificity*, which simply records the probability of a state (M_i's being *on*) existing conditional on a certain individual *a* being around: $P(on|a)$. In that case for IR to be possible, the state-specificity of a sufficient number of states must be sufficiently good. The 'sufficiently' and 'sufficient' here means: enough to explain that N's Fitness Contribution has been as high as to cause its fixation in the population of its possessors.

This strategy allows the existence of values of state-specificity well below 1; this will happen, for example, whenever the fitness-value of hits largely compensates for the fitness-value of misses. We may envisage states that have a content involving a concrete individual but such that the possessor of the state is perfectly unable to distinguish this individual from some other. Take, for example the lobster example examined in 3.9. It is conceivable that a certain lobster L_i develops a

state M_{ij} upon losing a fight with lobster L_j. But it may still be that the urine chemical signature UCS_j is shared by L_j with lobster L_k. In this situation, L_i will fail to distinguish L_j from L_k and nevertheless it is still the case that M_{ij} has as a content L_j, because it is L_j that has in its cluster the instantiation of the property *Being a token of* UCS_j that accounts for the existence of M_{ij}.

IR - CONTENT also explains why response-specificity is not necessary for IR: it is enough if the individual the mental state is about is situated in the right place of the causal structure that explains the existence of the contentful state; this is more often than not compatible with very imperfect response-specificity.

Tibbetts et al. (2008) offers an example in which there is no individual-specific response and no IR either: the discussion of meerkat alarm-calls in Schibler and Manser (2007). Meerkat alarm calls -the characteristic sound meerkats make when they perceive a predator, to warn other conspecifics- appear to be an individual-specific cue, at least in the informal, "for all intents and purposes" sense: the sound each meerkat makes is noticeably different from that of all other meerkats.

Meerkats responses, though, are not specific: meerkats flee no matter who, reliable or unreliable witness, issues the call. Tibbetts *et al.* propose this as evidence that without response-specificity there is no IR. IR - CONTENT provides a better explanation of the lack of IR in this case: is not that there is no response-specificity but, rather, that individual meerkats do not play the relevant role in the fitness-conduciveness of the mental mechanism that uses alarm calls as cues. This is a case in which the response of meerkats does not even accord to SOME: the response is *always* the same, no matter what. There is no need to appeal to the connection of individual-meerkat calls with the actual probability of danger; the connection of meerkat calls in general with such danger is doing all the work.

If they had played it, alarm calls would involve IR, even in the absence of response-specificity.

Experimental Setups.

A last point Tibbetts et al. make in Tibbetts et al. (2008) (and, to judge by the title of their letter, one that they see as particularly important) is that a proposal that leaves response-specificity out is difficult to test. Without response specificity we may not be able to ascertain whether a particular cue is individual-specific *for the agent* to whom we are ascribing IR. This may be an important practical problem, although there are ways to overcome it. The lobster case suggests one such a way: if there are only two possible responses, a way to test sensitivity to individually-specified cues is to assess the grouping of cues that prompt each one of the two responses. If the grouping is arbitrary from the perspective of the cue itself -that is, there are no features of the cue that ground the grouping-, this is evidence that it is the individually-grounded correlation between cue and appropriate response that drives the grouping. In the case of the meerkat, on the contrary, there is such a non-arbitrary grouping of the cues: *all* of them result in fleeing. This is evidence that meerkats are not doing IR.

In any event, practical problems such as this should have no bearing on the issue of what *constitutes* individual recognition. Researchers must simply think harder about experimental setups.

BIBLIOGRAPHY

Agar, N.: 1993, What Do Frogs Really Believe?, *Australasian Journal of Philosophy* **71**(1), 1–12. (Cited on pages 19, 40, and 106.)

Allaby, M. (ed.): 2009, *A Dictionary of Zoology*, Oxford Reference Online - Oxford University Press.
URL: *http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t8.e8577*
(Cited on page 10.)

Ayala, F.: 1970, Teleological Explanations in Evolutionary Biology, *Philosophy of Science* **37**(1), 1–15. (Cited on page 14.)

Bealer, G.: 2002, Modal Epistemology and the Rationalist Renaissance, *in* T. Gendler and J. Hawthorne (eds), *Conceivability and Possibility*, Oxford University Press, Oxford, pp. 71–125. (Cited on page 149.)

Bell, G.: 2008, *Selection. The Mechanism of Evolution*, Oxford University Press. (Cited on pages 158 and 170.)

Block, N.: 1986, Advertisement for a semantics for psychology, *Midwest Studies in Philosophy* **10**, 615–678. (Cited on page 144.)

Block, N.: 1998, *Routledge Encyclopedia of Philosophy*, Routledge, chapter Conceptual Role Semantics. (Cited on pages 143 and 144.)

Boghossian, P.: 1989, The Rule-Following Considerations, *Mind* **98**(392), 507–549. (Cited on page 127.)

Boyd, R.: 1988, *Moral Realism*, Cornell University Press, chapter How to be a Moral Realist, pp. 181–228. (Cited on pages 29, 65, and 199.)

Boyd, R.: 1991, Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds, *Philosophical Studies* **61**(1-2), 127–148. (Cited on page 30.)

Brigandt, I.: 2009, Natural Kinds in Evolution and Systematics: Metaphysical and Epistemological Considerations, *Acta Biotheoretica* **57**, 77–97. (Cited on page 30.)

Buller, D.: 1998, Etiological Theories of Function: A Geographical Survey, *Biology and Philosophy* **13**, 505–527. (Cited on page 15.)

Casullo, A.: forthcoming, Knowledge and modality, *Synthese* . (Cited on page 196.)

Chalmers, D.: 1996, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press. (Cited on page 193.)

Chalmers, D.: 2002, Does Conceivability Entail Possibility?, *in* T. Gendler and J. Hawthorne (eds), *Conceivability and Possibility*, Oxford University Press, pp. 145–200. (Cited on pages 149, 193, and 194.)

Chalmers, D.: 2009, The two-dimensional argument against materialism, *in* B. McLaughlin and A. Beckermann (eds), *The Oxford Handbook of Philosophy of Mind*, Oxford University Press. (Cited on page 193.)

Chalmers, D. J.: 2004, Epistemic two-dimensional semantics, *Philosophical Studies* **118**(1-2), 153–226. (Cited on page 194.)

Conee, E. and Feldman, R.: 1998, The generality problem for reliabilism, *Philosophical Studies* **89**(1), 1–29. (Cited on pages 181 and 182.)

Cooke, S. and Bliss, T.: 2006, Plasticity in the Human Central Nervous System, *Brain* **129**, 1659–1673. (Cited on page 88.)

Cummins, R.: 1975, Functional analysis, *Journal of Philosophy* **72**, 741–765. (Cited on page 14.)

Cummins, R.: 1991, *Meaning and Mental Representation*, The MIT Press. (Cited on page 36.)

Descartes, R.: 1641/1988, *Descartes: Selected Philosophical Writings*, Cambridge University Press. (Cited on page 188.)

Dretske, F.: 1981, *Knowledge and the Flow of Information*, The MIT Press. (Cited on page 8.)

Dretske, F.: 1988, *Explaining Behavior. Reasons in a World of Causes*, The MIT Press. (Cited on pages 5, 11, 24, 54, and 199.)

Enç, B.: 2002, Indeterminacy of function attributions, *in* A. Ariew, R. Cummins and M. Perlman (eds), *Functions: New Essays in the Philosophy of Psychology and Biology*, Oxford University Press, pp. 291–313. (Cited on page 24.)

Evans, G.: 1982, *The Varieties of Reference*, Oxford Clarendon Press. (Cited on page 133.)

Fodor, J.: 1986, Why Paramecia Don't Have Mental Representations, *in* U. French, P. (ed.), *Midwest Studies in Philosophy*, Vol. 10, University of Minnesota Press, pp. 3–23. En el contexto de la objeción del Liberalismo contra la teleosemántica. (Cited on page 73.)

Fodor, J.: 1990, *A Theory of Content and Other Essays*, The MIT Press. (Cited on pages 4, 19, 22, 106, and 107.)

Fodor, J.: 2008, *LOT 2*, Oxford Clarendon Press. (Cited on pages 107, 108, 140, and 201.)

Fodor, J. A. and Pylyshyn, Z. W.: 1988, Connectionism and cognitive architecture, *Cognition* **28**, 3–71. (Cited on page 107.)

Fodor, J., Garrett, M., Walker, E. and Parkes, C.: 1999, *Concepts: Core Readings*, The MIT Press, chapter Against Definitions. (Cited on page 136.)

Fodor, J. and Lepore, E.: 1992, *Holism: A Shopper's Guide*, Blackwell. (Cited on page 144.)

Fodor, J. and Piattelli-Palmarini, M.: 2010, *What Darwin Got Wrong*, Profile Books. (Cited on page 15.)

Forbes, G.: 1985, *The Metaphysics of Modality*, Oxford: Clarendon Press. (Cited on pages 167 and 173.)

Ghiselin, M.: 1974, A Radical Solucion to the Species Problem, *Systematic Zoology* **23**, 536–544. (Cited on page 95.)

Gillies, D.: 2000, Varieties of propensity, *British Journal for the Philosophy of Science* **51**(4). (Cited on pages 12 and 13.)

Godfrey-Smith, P.: 1994, A modern history theory of functions, *Noûs* **28**(3), 344–362. (Cited on page 17.)

Godfrey-Smith, P.: 1996, *Complexity and the Function of Mind in Nature*, Cambridge University Press. (Cited on pages 15, 54, and 199.)

Godfrey-Smith, P.: 2004a, Mental representation, naturalism, and teleosemantics, *in* D. Papineau and G. MacDonald (eds), *Teleosemantics: New Philosophical Essays*, Oxford University Press. (Cited on page 33.)

Godfrey-Smith, P.: 2004b, On folk psychology and mental representation, *in* H. Clapin, P. Staines and P. Slezak (eds), *Representation in Mind*, Perspectives in Cognitive Science, Elsevier, pp. 147–162. (Cited on page 33.)

Godfrey-Smith, P.: 2009a, *Darwinian Populations and Natural Selection*, Oxford University Press. (Cited on page 15.)

Godfrey-Smith, P.: 2009b, Representationalism reconsidered, *in* D. Murphy and M. Bishop (eds), *Stich and His Critics*, Philosophers and Their Critics, Wiley-Blackwell, pp. 30–45. (Cited on page 106.)

Goldman, A.: 2008, Reliabilism, *Stanford Encyclopedia of Philosophy*. (Cited on pages 181 and 185.)

Goldman, A. I.: 1976/2000, *Epistemology. An Anthology*, Blackwell, chapter What is Justified Belief?, pp. 340–353. (Cited on page 181.)

Goldman, A. I.: 1986, *Epistemology and Cognition*, Harvard University Press. (Cited on page 181.)

Grice, P.: 1957, Meaning, *Philosophical Review* **66**, 377–88. (Cited on page 7.)

Griffiths, P. E.: 1993, Functional analysis and proper functions, *British Journal for the Philosophy of Science* **44**(3), 409–422. (Cited on page 17.)

Hume, D.: 1740/1978, *A Treatise of Human Nature*, Oxford Clarendon Press. (Cited on page 189.)

Jenkins, C.: n.d., Concepts, experience and modal knowledge. (Cited on page 188.)

Karavanich, K. and Atema, J.: 1998, Individual Recognition and Memory in Lobster Dominance, *Animal Behavior* **56**(6), 1553–1560. (Cited on pages 97 and 216.)

Kingsbury, J.: 2006, A Proper Understanding of Millikan, *Acta Analytica* **21**(3), 23–40. (Cited on pages 14 and 80.)

Kment, B.: 2006, Counterfactuals and explanation, *Mind* **115**, 261–309. (Cited on page 170.)

Koch, C.: 1999, *Biophysics of Computation*, Oxford University Press. (Cited on page 88.)

Kripke, S.: 1980, *Naming and Necessity*, Harvard University Press. (Cited on pages 5, 167, and 193.)

Kripke, S.: 1982, *Wittgenstein on Rules and Private Language*, Harvard University Press. (Cited on page 127.)

Lansing, A.: 1999, *Endurance. Shackleton's Incredible Voyage*, Basic Books. (Cited on page 168.)

Laurence, S. and Margolis, E.: 1999, *Concepts. Core Readings*, The MIT Press, chapter Concepts and Cognitive Science. (Cited on pages 135 and 139.)

Lettvin, J. e. a.: 1959, What the Frog's Eye Tells the Frog's Brain, *Proceedings of the Institute of Radio Engineers* **49**, 1940–1951. (Cited on page 23.)

Lewis, D.: 1979, Counterfactual dependence and time's arrow, *Noûs* **13**(4), 455–476. (Cited on page 170.)

Loewer, B.: 1999, A guide to naturalizing semantics, *in* C. Hale, B.; Wright (ed.), *A Companion to the Philosophy of Language*, Blackwell: Oxford, pp. 108–126. (Cited on pages 4, 106, and 142.)

Macdonald, G. and Papineau, D.: 2006, *Teleosemantics*, Oxford University Press, chapter Prospects and Problems for Teleosemantics, pp. 1–22. (Cited on page 45.)

Machery, E.: 2009, *Doing Without Concepts*, Oxford University Press. (Cited on page 135.)

Mackie, P.: 1998, Identity, time, and necessity, *Proceedings of the Aristotelian Society* **98**(1), 59–78. (Cited on pages 167 and 173.)

Mackie, P.: 2006, *How Things Might Have Been*, Oxford University Press. (Cited on pages 167 and 174.)

Margolis, E. and Laurence, S. (eds): 1999, *Concepts: Core Readings*, The MIT Press. (Cited on page 135.)

Maynard-Smith, J.: 1999, *Evolutionary Genetics (2nd. Edition)*, Oxford University Press. (Cited on page 158.)

Mendola, J.: 2006, Papineau on etiological teleosemantics for beliefs, *Ratio* **19**(3), 305–320. (Cited on page 61.)

Millikan, R.: 1984, *Language, Thought and Other Biological Categories*, The MIT Press. (Cited on pages 11, 15, 30, 47, 64, 77, 78, 79, 81, 84, 85, 92, 119, 124, 125, 127, 133, 151, 169, and 199.)

Millikan, R.: 1989a, Biosemantics, *The Journal of Philosophy* **86**, 281–297. (Cited on page 48.)

Millikan, R.: 1989b, In Defense of Proper Functions, *Philosophy of Science* **56**(2), 288–302. (Cited on page 48.)

Millikan, R.: 1991, *Meaning in Mind. Fodor and his Critics*, Blackwell, chapter Speaking Up for Darwin, pp. 151–164. (Cited on page 46.)

Millikan, R.: 1993, *White Queen Psychology and Other Essays for Alice*, The MIT Press. Bradford Books. (Cited on pages 14, 45, 49, 127, 128, and 129.)

Millikan, R.: 1995, Pushmi-Pullyu Representations, *Philosophical Perspectives* **9, AI, Connectionism, and Philosophical Psychology**, 185–200. (Cited on page 93.)

Millikan, R.: 1998, A Common Structure for Concepts of Individuals, Stuffs, and Basic Kinds: More Mama, More Milk and More Mouse, *Behavioral and Brain Sciences* **22**(1), 55–65. (Cited on pages 91, 94, and 139.)

Millikan, R.: 2000, *On Clear and Confused Ideas*, Cambridge University Press. (Cited on pages 10, 30, 65, 94, and 110.)

Millikan, R.: 2002, *Functions: New Essays in the Philosophy of Psychology and Biology*, Oxford University Press, chapter Biofunctions: Two Paradigms, pp. 113–143. (Cited on pages 15, 47, 80, 81, 82, 83, 84, and 127.)

Millikan, R.: 2004, *Varieties of Meaning*, London: MIT Press. (Cited on pages viii, 13, 46, 50, 51, 123, 124, 133, and 199.)

Millikan, R.: 2006, Useless content, *in* G. Macdonald and D. Papineau (eds), *Teleosemantics*, Oxford University Press. (Cited on pages 131, 133, and 186.)

Millikan, R.: 2007, An Input Condition for Teleosemantics? A reply to Shea (and Godfrey-Smith), *Philosophy and Phenomenological Research* **75**(2). (Cited on pages 12, 50, 55, and 57.)

Millikan, R.: 2009, *The Oxford Handbook of Philosophy of Mind*, Oxford University Press, chapter Biosemantics, pp. 394–406. (Cited on pages 46, 49, 69, and 199.)

Millikan, R. G.: 1990, Truth, rules, hoverflies, and the Kripke-Wittgenstein paradox, *Philosophical Review* **99**(3), 323–53. (Cited on page 49.)

Millikan, R. G.: 1997, Troubles with Wagner's reading of Millikan, *Philosophical Studies* **86**(1), 93–96. (Cited on page 80.)

Mossio, M., Saborido, C. and Moreno, A.: forthcoming, An organizational account of biological functions, *British Journal for the Philosophy of Science* . (Cited on page 14.)

Müller, J., Eggert, A.-K. and Elsner, T.: 2003, Nestmate Recognition in Burying Beetles: the "Breeder's Badge" As a Cue Used by Females to Distinguish Their Mates from Male Intruders, *Behavioral Ecology* **14**(2), 212–220. (Cited on page 215.)

Neander, K.: 1991, Functions as Selected Effects: The Conceptual Analyst's Defence, *Philosophy of Science* **58**, 168–184. (Cited on pages 14 and 44.)

Neander, K.: 1995, Misrepresenting & Malfunctioning, *Philosophical Studies* **79**, 109–141. (Cited on pages 15, 19, 22, 49, and 67.)

Papineau, D.: 1987, *Reality and Representation*, Basil Blackwell. (Cited on pages 58 and 77.)

Papineau, D.: 1993, *Philosophical Naturalism*, Basil Blackwell. (Cited on pages 58, 60, and 61.)

Papineau, D.: 1998, Teleosemantics and Indeterminacy, *Australasian Journal of Philosophy* **76**(1), 1–14. (Cited on pages 58, 59, and 60.)

Papineau, D.: 2001, The Status of Teleosemantics, or How to Stop Worrying About Swampman, *Australasian Journal of Philosophy* **79**(2), 279–89. (Cited on page 62.)

Papineau, D.: 2006, *The Oxford Handbook of Philosophy of Language*, Oxford: OUP, chapter Naturalist Theories of Meaning, pp. 175–188. (Cited on pages 4 and 77.)

Papineau, D.: 2007, *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*, Oxford University Press, chapter Phenomenal and Perceptual Concepts, pp. 111–144. (Cited on pages 149 and 150.)

Parfit, D.: 1984, *Reasons and Persons*, Oxford: Oxford University Press. (Cited on page 92.)

Peacocke, C.: 1992, *A Study of Concepts*, The MIT Press. (Cited on pages 134 and 186.)

Peacocke, C.: 1999, *Being Known*, Oxford University Press. (Cited on page 151.)

Peacocke, C.: 2008, *Truly Understood*, Oxford University Press. (Cited on page 140.)

Pérez-Otero, M.: 1997, La concepción ramificacionista de la modalidad, *Contextos* **XV**, 135–152. (Cited on pages 167 and 173.)

Pietroski, P.: 1992, Intentionality and Teleological Error, *Pacific Philosophical Quarterly* **73**, 267–282. (Cited on page 68.)

Preston, B.: 1998, Why Is a Wing Like a Spoon? A Pluralist Theory of Function, *The Journal of Philosophy* **95**(5), 215–254. (Cited on page 85.)

Price, C.: 1998, Determinate Functions, *Noûs* **32**(1), 54–75. (Cited on pages 15, 22, 24, 42, 43, and 44.)

Putnam, H.: 1975, The meaning of 'meaning', *Minnesota Studies in the Philosophy of Science* **7**, 131–193. (Cited on page 5.)

Quine, W. V. O.: 1953, Two dogmas of empiricism, *From a Logical Point of View*, Harvard University Press. (Cited on page 136.)

Quine, W. V. O.: 1960, *Word & Object*, The MIT Press. (Cited on page 91.)

Revonsuo, A.: 2009, The binding problem, *The Oxford Companion to Consciousness*, Oxford University Press, pp. 101–105. (Cited on page 163.)

Richards, R.: 2008, *The Oxford Handbook of Philosophy of Biology*, Oxford University Press, chapter Species and Taxonomy, pp. 161–188. (Cited on page 30.)

Roca-Royes, S.: forth., Conceivability and de re modal knowledge, *Noûs* . (Cited on page 195.)

Rowlands, M.: 1997, Teleological Semantics, *Mind* **106**(422), 279–303. (Cited on pages 19 and 22.)

Rupert, R.: 1999, Mental Representations and Millikan's Theory of Intentional Content: Does Biology Chase Causality?, *The Southern Journal of Philosophy* **37**(3), 113–140. (Cited on page 134.)

Rupert, R. D.: 2008, Causal theories of mental content, *Philosophy Compass* **3**, 353–380. (Cited on pages viii, 4, and 9.)

Ryder, D.: 2004, Sinbad neurosemantics: A theory of mental representation, *Mind & Language* **19**(2), 211–240. (Cited on pages viii, 24, 30, and 142.)

Ryder, D.: 2006, On thinking of kinds, *in* G. Macdonald and D. Papineau (eds), *Teleosemantics*, Oxford University Press, pp. 1–22. (Cited on page 142.)

Sauchelli: forthcoming, Concrete possible worlds and counterfactual conditionals: Lewis versus williamson on modal knowledge, *Synthese* . (Cited on page 169.)

Schibler, F. and Manser, M.: 2007, The Irrelevance of Individual Discrimination in Meerkat Alarm Calls, *Animal Behaviour* **74**(5), 1259–1268. (Cited on page 218.)

Schiffer, S.: 1996, Contextualist solutions to scepticism, *Proceedings of the Aristotelian Society* **96**, 317–333. (Cited on page 53.)

Schroeder, T.: 2004, New norms for teleosemantics, *in* H. Clapin (ed.), *Representation in Mind*, Elsevier. (Cited on page 14.)

Searle, J.: 1958, Proper names, *Mind* **67**, 166–173. (Cited on page 137.)

Shea, N.: 2004, *On Millikan*, Wadsworth. (Cited on page 80.)

Shea, N.: 2007, Consumers Need Information: Supplementing Teleosemantics with an Input Condition, *Philosophy and Phenomenological Research* **75**(2), 404–435. (Cited on pages viii, 12, 39, 50, 54, 56, and 57.)

Sherman, P., Reeve, H. and Pfennig, D.: 1997, *Behavioural Ecology, 4th Edition*, Oxford: Blackwell Science, chapter Recognition Systems, pp. 69–96. (Cited on page 215.)

Sherry, R. A. and Galen, C.: 1998, The mechanism of floral heliotropism in the snow buttercup, *Plant, Cell and Environment* **21**, 983–993. (Cited on page 93.)

Shoemaker, S.: 1998, Causal and Metaphysical Necessity, *Pacific Philosophical Quarterly* **79**, 59–77. (Cited on pages 66 and 173.)

Stampe, D.: 1977, Towards a Causal Theory of Linguistic Representation, *Midwest Studies in Philosophy* **2**, 42–63. (Cited on pages 6, 7, 13, and 24.)

Steiger, S. and Müller, J.: 2008, 'True' and 'Untrue' Individual Recognition: Suggestion of a Less Restrictive Definition, *Trends in Ecology and Evolution* **23**(7), 355. (Cited on page 215.)

Sterelny, K.: 1990, *The Representational Theory of Mind: An Introduction*, Oxford University Press. (Cited on pages 9 and 67.)

Stuart-Fox, D. and Moussalli, A.: 2008, Selection for social signalling drives the evolution of chameleon colour change, *PLoS Biology* **6**(1), e25. (Cited on page 78.)

Szabó, Z.: 2008, *The Stanford Encyclopedia of Philosophy (Winter 2008 edition)*, chapter Compositionality.
URL: *http://plato.stanford.edu/archives/win2008/entries/compositionality/* (Cited on page 117.)

Tibbetts, E. and Dale, J.: 2008, Individual Recognition: It is Good to be Different, *Trends in Ecology and Evolution* **22**(10), 529–537. (Cited on pages 215 and 216.)

Tibbetts, E., Sheehan, M. and Dale, J.: 2008, A Testable Definition of Individual Recognition, *Trends in Ecology and Evolution* **23**(7), 356. (Cited on pages 215, 216, and 218.)

Tye, M.: 2000, *Consciousness, Color and Content*, The MIT Press. Bradford Books. (Cited on page 68.)

Van Inwagen, P.: 1998, Modal Epistemology, *Philosophical Studies* **92**, 67–84. (Cited on pages 168, 191, and 203.)

Weiner, M. and Belnap, N.: 2006, How causal probabilities may fit into our objectively indeterministic world, *Synthese* **149**, 1–36. (Cited on page 13.)

Williamson, T.: 1996, Knowing and asserting, *Philosophical Review* **105**(4), 489–523. (Cited on page 103.)

Williamson, T.: 2000, *Knowledge and Its Limits*, Oxford University Press. (Cited on pages 103 and 180.)

Williamson, T.: 2008, *The Philosophy of Philosophy*, Wiley-Blackwell. (Cited on page 170.)

Wilson, R., Barker, M. and Brigandt, I.: forthcoming, When Traditional Essentialism Fails: Biological Natural Kinds, *Philosophical Topics* **35**(1-2). (Cited on pages 30 and 94.)

Wittgenstein, L.: 1953/1973, *Philosophical Investigations*, Prentice Hall. (Cited on page 5.)

Worley, S.: 2003, Conceivability, Possibility, and Physicalism, *Analysis* **63**, 15–23. (Cited on page 195.)

Wright, L.: 1973/1994, *Conceptual Issues in Evolutionary Biology*, The MIT Press. Bradford Books, chapter Functions, pp. 27–48. (Cited on page 14.)

Yablo, S.: 1993, Is Conceivability a Guide to Possibility?, *Philosophy and Phenomenological Research* **53**(1), 1–42. (Cited on pages 149, 188, 189, and 191.)

Zawidzki, T.: 2003, Mythological Content: A Problem for Millikan's Teleosemantics, *Philosophical Psychology* . (Cited on page 106.)

COLOPHON

This thesis was typeset in LyX, available from www.lyx.org. The typographic style is available for LaTeX via CTAN as "`classicthesis`". A classicthesis port for LyX is available from www.soundsorange.net.

Some figures were created using the Biggles module for Python (biggles.sourceforge.net), others using SciLab (www.scilab.org) and the rest using OpenOffice (www.openoffice.org).

Final Version as of 4th May 2010 at 16:21.