Universitat de Girona

A NATURALISTIC THEORY OF INTENTIONAL CONTENT

Marc ARTIGA GALINDO

Dipòsit legal: Gi. 1385-2013
http://hdl.handle.net/10803/123436

Universitat de Girona

Doctoral Thesis

A NATURALISTIC THEORY OF INTENTIONAL CONTENT

MARC ARTIGA GALINDO

2013

Universitat de Girona

Tesi Doctoral
Doctor per la Universitat de Girona

Titol
A Naturalistic Theory of Intentional Content

Autor
Marc Artiga Galindo

Tutor
David Pineda Oliva

Programa
Programa de Doctorat de Ciències Humanes i de la Cultura

Universitat de Girona
2013

Als meus pares i a la meva germana

A la Vero

DECLARACIÓ

El Dr. David Pineda Oliva, de la Universitat de Girona,

CERTIFICO: Que aquest treball , titulat "A Naturalistic Theory of Intentional Content", que presenta Marc Artiga Galindo per a l'obtenció del títol de doctor, ha estat realitzat sota la meva direcció i que compleix els requeriments per poder optar a Menció Internacional.

I, perquè així consti i tingui els efectes oportuns, signo aquest document.

Girona,

ACKNOWLEDGMENTS

Over the years, so many philosophers and friends have helped me develop the ideas contained in this dissertation, that it is impossible to mention them all. I feel that I have been extremely lucky.

First of all, I would like to thank my research group, Logos, for providing me with an enviable philosophical environment. I really think I have learned to philosophize at Logos, so if anything of what I have written in this dissertation makes sense, this is to a great extent thanks to the Logos community. Among all logosians, I would like to specially thank those who have read parts of this thesis or who have helped me to form some of the ideas contained on it: Miguel Àngel Sebastian, Marta Jorba, Mireia Lopez and Manolo Martinez. I am also greteful to the Universitat de Girona for hosting me, and providing me with the material and financial support I needed.

During the last five years of research, I have been lucky enough to have the chance of traveling around the globe and visit some of the most exciting places in the world. In every place I have been, I have met outstanding philosophers and lovely people. When I was in Connecticut (US), I met Jesse Mulder, who among other things, taught me English. Donald Shankweiler, BJ Strawser, Marcus Rossberg and Steven Todd helped me very much both emotionaly and intellectually.

I also found a very stimulating and enjoyable atmosphere at the Australasian National University. I would like to thank Kim Sterelny for accepting my visit and allowing me to participate in the activities of one of the best australasian universities. Among all the people I met and all the things I learned, perhaps the greatest lesson was given by Class Weber, Leon Leontyev and Alma Barner, who taught me that, if one puts enough effort, any obstacle can be overcome.

My last stay abroad was at King's College London, where David Papineau was extremely helpful and patient with me, and I have but good words and thoughts for him. My alter ego in London was Benoit Conti, who taught me many things about concepts and life. I really would like to thank them all for their lessons.

Ruth Millikan deserves a particular acknowledgement. She accepted my visit to Connecticut, without knowing me or having any reference whatsoever. She has encouraged me every time I have needed help and even offered me to live at her place for some months. She has shaped my ideas more than anyone else and has been the main source of philosophical inspiration. To a great extent, this dissertation is a bit of *Millikanianism*. Anyone reading this thesis will find continous references to her work and extensive discussions of her main theses. So, even if one can find some places of disagreement, this dissertation can be regarded as a deep effort of discussing and completing the theory that she has developped over the years. Thank you.

My debt to Manolo Martínez is also a special one. Probably without realizing it, he has been a contiuous source of stimulating ideas for me. He is an amazing teleosemanticist and a better person, and having him around has surely made this dissertation much better than it would have otherwise been.

I also want to thank David Pineda for supervising my thesis and helping me over the years. He accepted being my supervisor without hardly knowing me, and allowed me to make long stays abroad. He has meticulously read this thesis several times and made insightful cricitisms to many of my (old and new) ideas, which have clearly improved this thesis.

Furthermore, I profited from discussion of audiences in Valencia, San Sebastian, Santiago de Compostela, Barcelona, Girona, Lisbon, Canberra, Sydney, Geneve, La Laguna and Évora. The questions, comments, criticisms and long discussions with dozens of philosopers in talks, dinners, lunchs and parties have contributed to many of the ideas contained in the dissertation. I would like to thank them as well.

Finalment, volia agraïr els meus pares, a la meva germana i a la Vero tot el suport durant aquests anys. Per completar un doctorat, tan important és el suport intel·lectual com el suport anímic. Se'm fa molt difícil enumerar tot allò en què m'han ajudat, i qualsevol cosa que digui es quedarà curta. Gràcies a ells sóc qui sóc, i he fet el que fet. Sense el seu ajut incondicional, els seus ànims, l'amor i la tranquilitat que m'han transmès durant anys, res d'això hauria estat possible. Moltes gràcies per ajudar-me en els moments difícils (els nervis, les pors, els dubtes,...) i també per compartir amb mi les alegries i les petites victòries. Gràcies per tot aquest suport emocional que només vosaltres em podieu donar. Aquesta tesi us la vull dedicar a vosaltres.

LIST OF PUBLICATIONS

- Artiga, M. (Forthcoming) Teleosemantics and Pushmi-Pullyu Representations, *Erkenntnis*
- Artiga, M. (Forthcoming) Reliable Misrepresentation and Teleosemantics, *Disputatio*
- Artiga, M. (2012) The Singular Thought Strategy and the Contents of Perception, C. Martínez Vidal, Falguera J.L., Sagüillo, J.M., Verdejo, V.M., Pereira-Fariña, M., (Eds) *Proceedings of the VII Conference of the Spanish Society for Logic, Methodology and Philosphy of Science, Santiago de Compostela* (Spain): USC Press, pp. 114-121
- Artiga, M. (2011) Several Misuses of Sober's Selection for/Selection of distinction. *Topoi, An International Review of Philosophy,* Vol. 30 (2): p. 181-193
- Artiga, M. (2011) Re-Organizing Organizational Accounts of Function. *Applied Ontology,* vol. 6, p. 105-124
- Artiga, M. (2010) Learning and Selection Processes. *Theoria, An International Journal for Theory, History and Foundations for Science,* Vol 25(2): Nº68, pp. 197- 210
- Artiga, M. (2010) Teleosemantics and the Indeterminacy Problem. Jaume, A., Liz, M., Perez Chico, D., Ponte, M., Vázquez, M. (Eds.) *Proceedings of the VI Conference of the Spanish Society for Analytic Philosophy,* Tenerife, Universidad de la Laguna, pp.29-30
- Artiga, M. (2009) Against Original Intentionality. Alcolea, J. Iranzo, V, Sánchez, A., and Valor, J. (Eds.) *Proceedings of the VI Conference of the Spanish Society of Logic, Methodology and Philosophy of Science,* València, Universitat de València, pp. 145-151

LIST OF FIGURES

ABSTRACT

This dissertation develops a naturalistic theory of intentional content. In the first part, a teleosemantic theory of content is defended, based on a precise definiton of the notions of function and sender-receiver system, and several objections are addressed. In the second part, the teleosemantic theory developed in the first chapters is applied to perceptual representations and concepts.

RESUM

Aquesta tesi desenvolupa una teoria naturalista del contingut intencional. A la primera part es defensa una teoria teleosemàntica del contingut, basada en una definició precisa de les nocions de funció i sistema emisor-receptor, i es responen una sèrie d'objeccions. A la segona part, la teoria teleosemànica desenvolupada en els primers capítols és aplicada a les representacions perceptives i als conceptes.

RESUMEN

Esta tesis deserrolla una teoría naturalista del contenido intencional. En la primera parte se defiende una teoría teleosemántica del contenido, basada en una definicion precisa de las nociones de función y sistema emisor-receptor, y se responen una serie de objeciones. En la segunda parte, la teoría teleosemánica deserrolada en los primeros capítulos es aplicada a las representaciones perceptivas y a los conceptos.

CONTENTS

Part I

TELEOSEMANTICS

1 INTRODUCTION

One of the most perplexing facts about our world is that some states are *about* other states. Linguistic expressions, brain events and states of artifacts like thermometers or gas gauges have the capacity to be *about* other things. These devices differ, for instance, from stones or chairs which (usually) are not about anything. This is a phenomenon that has fascinated humanity for centuries and that still awaits a satisfactory explanation.

Historically, different terms have been used in order to talk about this phenomenon. Some verbs employed to that end are 'to mean', 'to refer', 'to signify', or 'to denote'. The most general expression we have nowadays, whose use in modern times we owe to Brentano, is 'intentionality'. Language, thought and many other phenomena are said to exhibit intentionality, i.e. the capacity to be about other things. This dissertation is intended as an analysis of what intentionality is and how it might have appeared in the natural world.

Intentional states are puzzling entities. Arguably, aboutness is a sort of relation; but it has certain striking features that other relations lack. For instance, while a relation like *being next to* presupposes the existence of the relata (A cannot be next to B if either A or B do not exist), it seems that a state can be about another state even if the latter does not exist. I can think about Wittgenstein's son or about unicorns, even if they have never existed.

A second conspicuous feature of intentional states is that it is mysterious what process endows particular states with intentionality. Whereas we all know what makes it the case that A and B instantiate the relation *being next to* (occupying certain relative location), it is not obvious at all what grounds the fact that A is about B.

Trying to understand better these and other perplexing features of intentionality has been the main motivation of much philosophical work. There are many ways in which we can try to clarify the notion of intentionality. This dissertation assumes a particular strand in this tradition and seeks to push it a bit further. My main purpose is to *naturalize* the phenomenon of intentionality.

Thus, in order to understand better the goal of this dissertation and the tradition I am relying on, we need to define in more detail several notions.

1.1 BASIC NOTIONS

1.1.1 *Naturalism (*sive *Physicalism)*

Naturalism is a very controversial notion (Papineau, 2007). Some people think it is the same as physicalism, while others think these are different claims. In this dissertation I will use both expressions interchangeably; that is, by 'naturalism' I mean physicalism (Papineau, 1993, 2007). Needless to say, that is not very explanatory unless the notion of 'pyshicalism' is spelled out.

There are two important aspects of physicalism (or naturalism) that need to be settled here, what Stoljar (2010) calls the *interpretation question* and the *truth question.* The interpretation question asks: what does 'physicalism' mean? In contrast, the truth question asks: why should we believe physicalism is true? Let me briefly go over these two questions.

1.1.1.1 *Interpretation Question*

What is physicalism? As a first approximation, I take physicalism to be a thesis about the supervenience of a set of properties on physical properties.

In this context, the expression 'supervenience' is a term of art. A property A supervenes on a property B iff there can be no difference in A-properties without there being a difference in B-properties. In other words, if we fix all instantiations of B-properties at a given world *w*, then we thereby fix all instantiations of A-properties at that world *w.*

Now, we can distinguish *local* physicalism from *global* physicalism. Global physicalism is the claim that everything supervenes on, or is necessitated by, the physical (Stoljar, 2009, p.56). Global physicalism claims that once you specify the physical properties of a world, then the rest of properties are also fixed. That is, there can be no difference in any properties (biological, geological, mental,...) without a difference in physical properties. Of course, this is only a first approximation; several problems show that this simple definition of global physicalism is probably wrong or, at least, incomplete (e.g. Stoljar, 2009, 2010; Tye, 2009a). For instance, if God exists, then there is a necessary being that exists and remains identical in all possible worlds. As a result, his existence does not falsify the claim that there can be no difference in supervenient properties without a difference in the supervenience base. So, if God exists, this formulation of Global physicalism is true, but intuitively the existence of God is a scenario that should falsify physicalism (Jackson, 1998). Similar counterexamples have been presented (Hawthorne, 2002; Stoljar, 2009).

Fortunately for our purposes, one can be a physicalist about a certain domain, without having to be a global physicalist. One could, for instance, hold that mental properties supervene o physical properties and, at the same time, hold that there is an élain vital that does not supervene on the physical. This is a form of what I call *local* physicalism, i.e. physicalism about a particular domain.

In this dissertation, we will only be concerned with semantic properties and semantic facts.[1] Intentional properties are semantic properties. So in order to frame the naturalistic project of this thesis, we only need to define a local physicalism concerning semantic properties or facts. Hopefully, that will leave some of the problems of global physicalism aside.

Thus, as a first approximation, the claim I would like to lend support to is the following:

LOCAL PHYSICALISM Semantic facts supervene on physical facts

As a first approximation, LOCAL PHYSICALISM means that there can be no difference in semantic facts without a difference in physical facts. Now, there are four issues that need to be clarified; first of all, we need

1 I take it that any formulation of physicalism in terms of properties can be translated into a definition in terms of facts and viceversa (Stoljar, 2010, p.41).

to specify the set of semantic facts that will be addressed; the second question concerns the modality involved in Local physicalism; thirdly, the notion of *supervenience* and finally the concept of *physical* employed.

SEMANTIC FACTS The world is pervaded with semantic facts. Every sentence on a book, any utterance of a speaker in any language, any communication signal among animals, any state of a barometer, any screen image on a TV constitutes a semantic fact. Explaining how all these facts supervene on physical facts is an enormous task that I cannot attempt to fully address in this dissertation. Instead, I will focus my attention on a set of semantic facts that have a *privileged* status: perceptual and conceptual states. In the first part of the dissertation I will explain what semantic states are and how they originate in simple organisms. In the second half I will show how this framework can be extended to perceptual mechanisms and human thoughts.

In a sense, we can say that, among all semantic facts, in this thesis we will be concerned with a particular set of 'mental facts' (assuming a broad interpretation of 'mental'), which constitute a privileged category. But, one can reasonably ask, why should we think these mental states are special? Why are perceptual and conceptual representations more interesting for a naturalistic project than gas gauges and images on TV screens? This particular set of semantic facts is privileged because there are good reasons for thinking that, if we succeed in providing a naturalistic explanation of these mental facts, we will probably be able to somehow extend this approach to the rest of semantic facts. There are two ways people have usually thought that expansion could be accomplished:

- First, it might be that the recipe I am about to offer for the naturalization of these privileged set of mental facts indeed applies to *all* semantic facts. For instance, I will sketch how semantic facts concerning communication signals among many animals can be easily accounted for with the naturalistic framework described here. Similarly, Millikan (1999, 2005) has attempted to extend her naturalistic account of representational states in simple organisms to human language and artifacts. So perhaps language and artifacts acquire semantic properties in exactly the same way simple representations do.

- There is a different way a naturalistic account of the representational properties of mentality can set the ground for an explanation of many other semantic properties: a standard assumption in philosophy is that the intentional properties of states in language and artifacts somehow *derive* from the semantic properties of the mind (Grice, 1989; Searle, 1983). That is, it is held that the semantic properties of, say, linguistic expressions supervene on mental facts. In a nutshell, the idea is that once you fix this privileged set of mental states at w, all the semantic properties of language and artifacts at w are also fixed.

Now, if any of these ways of extending a naturalistic account of mental states succeeded, that would vindicate the idea that the set of semantic facts I will be concerned here is, in some sense, privileged. Either if the rest of semantic facts can be accounted for using the same tools I devise for the privileged set or if they just supervene on the semantic facts I do account for, the naturalistic theory of mentality offered here

will provide the fundamental aspects of a naturalistic treatment of all semantic facts. This is the main reason I will focus on the following set of mental facts: the semantic properties of cognitively unsophisticated organisms, perceptual states and human thoughts. If I manage to account for the intentional properties of simple and complex minds, its extension to other domains that contain intentionality will not be hard to come by.

MODALITY Supervenience is a relation with modal import; it claims that there *can be* no difference of one kind of entity without a difference in the other. What kind of modality is involved in LOCAL PHYSICALISM?

A first obvious candidate is *nomological* supervenience: A-facts *nomologically* supervene on B-facts iff there is no nomologically possible world[2] where there is a difference in A without a difference in B. Unfortunately, it is usually held that *nomological* supervenience is insufficient for physicalism. For one thing, dualists can hold that there are contingent laws between physical and mental properties (Chalmers, 1996). So dualism is compatible with the nomological supervenience of mental facts on physical facts. Therefore, physicalism needs to be cashed out in terms of the stronger relation of *metaphysical* supervenience. Hence, LOCAL PHYSICALISM is the claim that semantic facts metaphysically supervene on physical facts, i.e. that there is no metaphysically possible world where there is a difference in semantic facts without a difference in physical facts.[3]

SUPERVENIENCE Secondly, one might worry that supervenience is too weak a relation for defining any form of physicalism, since the supervenience of semantic facts on physical facts is compatible with some forms of non-reductionism. For instance, epiphenomenalism (the view that semantic facts are caused by physical events in the brain, but have no effects upon any physical events) seems to be compatible with it.

Two things can be said to dispel this worry. First, notice that epiphenomenalism (or any other form of non-reductionism) is compatible with LOCAL PHYSICALISM only if it is granted that there is a necessary connection between semantic facts and physical facts. However, many people deny that there can be a necessary connection between metaphysically distinct entities (see Lewis, 1994; Armstrong 1997). So it is not entirely clear that it is plausible or even coherent to hold that semantic facts metaphysically supervene on physical facts and that, nevertheless, they are metaphysically distinct (for a discussion, see Stoljar, 2010 ch. 8).

Secondly, I hope this dissertation makes clear that the view defended here is not epiphenomenalist, bur reductionist. What I intend to show is that when we rightly attribute a semantic property to a given state, this attribution is true in virtue of some (complex) non-semantic process. I will argue that, in the same sense that being bald is nothing more than having a certain number of hairs, having a semantic property is nothing more than being in a state within a certain causal system. Hence, even if some forms of epiphenomenalism are compatible with

2 The set of nomologically possible worlds includes all the possible worlds with the same natural laws as the actual world.

3 Some people claim that dualism is compatible with there being a metaphysical supervenience between physical facts and semantic facts, but I think this claim is far less convincing (see below).

Local physicalism, I hope that the way I defend it readily shows that this dissertation presents a reduction of semantic phenomena to non-semantic facts (see below).

the physical It is well known that any attempt to define 'physical facts' faces Hempel's dilemma.[4] In a nutshell, the problem is the following: either 'physical facts' refers to facts posited by actual physics or it refers to facts posited by a future ideal and complete physical theory. Now, if we take the first horn of the dilemma and assume 'physical facts' refers to facts posited by *current* physics, physicalism is probably false; the history of science teaches us that current physics is likely to evolve in such a way as to render false most of the claims it actually holds (Chakravartty, 2011; Papineau, 1987). So, if we take this horn of the dilemma, physicalism is likely to be false. But the other horn is not better; if we attempt to define 'physical facts' by appealing to a future ideal and complete physical theory, physicalism becomes extremely uninformative; since we have no idea how this future and complete physics will look like, we cannot fully grasp what physicalism really means or even whether it is an account worth defending.

I think that this is a serious problem for any formulation of physicalism. Nevertheless, I think in this dissertation we can bypass this difficulty, because I will not attempt to reduce semantic facts to facts posited by (actual or future) physics. For one thing, I am not competent to do so; but, more importantly, at the current stage of research it is probably impossible to provide the physical details of the supervenience base of semantic facts.

The goal of this dissertation is much more modest (and, nevertheless, I hope interesting enough). I will try to reduce semantic facts to a class of facts that I will call 'φ- facts'. φ- facts are facts posited by current science that are very likely to supervene on physical facts (so they might include, but are not restricted to, physical facts). For instance, φ- facts include geological facts, chemical facts and (crucially) certain physiological and biological facts. Hence, I will show how semantic phenomena reduce to the presence of certain causal relations, systems, neurons, chemical compounds... that is, on entities that are very likely to supervene on physical facts, no matter whether *physical facts* are defined in terms of current or future physics. Of course, if semantic facts reduce to φ- facts and φ- facts supervene on physical facts, then semantic facts supervene on physical facts. This dissertation is an attempt to show that the first antecedent is true.[5]

Notice that reducing semantic facts to φ- facts is not an easy task; even more, it could be reasonably argued that this is the philosophically interesting question (at least, if we focus on the phenomenon of intentionality). Once semantic phenomena are accounted for in terms of chemistry, causal relations, neurons and so on, one might

4 Here is a quote where Hempel makes the dilemma explicit: "I would add that the physicalist claim that the language of physics can serve as a unitary language of science is inherently obscure: the language of what physics is meant? Surely not that of, say, eighteenth-century physics; for it contains terms like 'caloric fluid,' whose use is governed by theoretical assumptions now thought false. Nor can the language of contemporary physics claim the role of unitary language, since it will no doubt undergo further changes too. The thesis of physicalism would seem to require a language in which a true theory of all physical phenomena can be formulated. But it is quite unclear what is to be understood here by a physical phenomenon, especially in the context of a doctrine that has taken a decidedly linguistic turn." (Hempel, 1980, p. 194–5)

5 Offering a reduction of semantic facts to φ- facts is similar to what Horgan (1994) calls 'providing a tractable specification' of semantic facts.

plausibly conclude that (using a popular expression from the debate on consciousness) the *hard problem* of intentionality has been solved.

So, this dissertation is intended to argue for what I will call Local φ−physicalism:

Local φ−physicalism Semantic facts metaphysically supervene on (indeed, reduce to) φ- facts.

And I will be assuming the following claim:

Background Physicalism φ- facts metaphysically supervene on physical facts.

Where φ- facts are (1) facts posited by current science (such as causal relations, cells or organisms), (2) which are likely to be accepted by future science. I take it that it is extremely likely that φ- facts supervene on physical facts (whether they are defined by current or future physics).

To conclude this preliminary discussion of naturalism, let me consider a final objection; one might argue that Hempel's dilemma, which threatened Local physicalism, recurs in Local φ−physicalism. After all, φ- facts are facts posited by special sciences (geology, chemistry, biology,..). So one might ask whether φ- facts are facts postulated by current special sciences or by a future and complete development of these sciences. However, in contrast to Hempel's dilemma applied to Local physicalism, I think at this point we can confidently take the first horn of the dilemma; while it is very likely that physics will suffer major changes and hence that future physical theories will deny that the most fundamental properties are the ones postulated by current physics, it is hard to imagine that geology will ever deny that mountains or calcareous rocks exist, or biology will ever deny that organisms, neurons or even natural selection exists. Of course, these claims may still turn out to be false. But I take that to be a risk of any naturalistic endeavor. Any reasonable and well-informed naturalistic theory utterly depends on the validity of certain empirical claims; the key question is whether these empirical statements are reliable. And I think those facts I will rely on are extremely well-established in the respective scientific fields. Therefore, the kind of facts I will appeal to in this project are such that worries concerning the above dilemma can be, if not entirely avoided, at least reasonably left aside.

1.1.1.2 *Truth Question*

So far I have only been trying to provide a plausible and informative definition of (semantic) naturalism, but I have not offered any argument for why we should think such a form of naturalism is indeed true. Why should we believe semantic properties supervene on φ- facts?

Two arguments motivate this approach. First, the recent and surprising advance of science has been able to reduce many of the properties we are familiar with to more basic ones. Classical examples are the reduction of heat to mean kinetic energy, or the reduction of chemical properties to the properties of atoms. Certainly, semantic properties are special in many respects, but the striking progress of science in a vast range of different fields gives prima facie plausibility to any naturalistic project.

Secondly, even if one is not convinced by scientific progress or has independent reasons for thinking naturalism should not be the default

assumption, I think Occamist reasons suggest that *if* such a reductionist account is possible, *then* it is a preferable hypothesis over less reductionist accounts. In other words, if we can account for semantic facts in terms of φ- facts and no property of semantic facts is left out, I think this approach should strongly be preferred over another theory according to which semantic entities are fundamental or non-reducible. And that is precisely the project I intend to pursue in this dissertation. Hence, the whole dissertation can be regarded as a (long) argument in favor of naturalism concerning intentional properties. If I succeed in the reduction of semantic facts to φ- facts, it should be considered a strong argument in favor of naturalism, even in case one is willing to deny the initial plausibility of such a position.

1.1.2 *Representation*

But what it is to account for intentional properties in terms of φ- facts? In this work I will adopt a common strategy among people working on naturalistic approaches to intentionality: I will introduce the *semi*-technical notion of *representation* and during the dissertation I will be showing two things. First, I will extensively argue that representational facts supervene on (and, indeed, reduce to) φ- facts. Secondly, I will show that intentional states just *are* representational states. The strategy, hence, is to analyze intentional properties in terms of representational properties, and then to seek to naturalize the phenomenon of representation. I think there are three good reasons for using the notion of representation in that project.

First, the notion of 'intentional state' is ambiguous between a weaker and a stronger reading. Often, intentional states are defined as states that are about other states, as states that have truth or satisfaction conditions (see above). But other times, intentional states are restricted to propositional attitudes such as beliefs or desires. A consequence of that ambiguity is that, for instance, the claim that a set of neurons in early visual processing is an intentional state sounds obviously wrong to many people (in fact, it sounds like a *conceptual* mistake), whereas asserting that they represent the presence of an edge is far more palatable. The notion of representation might help us to bypass this ambiguity.

A second related advantage is that 'representation' is very much used in the cognitive sciences, to the extent that for instance Sternberg (2009) claims it is the unifying concept for the different disciplines that constitute the cognitive sciences. The notion of representation used here is intended to comprehend (but need not be restricted to) the notion used in cognitive science. In other words, representational states in cognitive science are representational states in the sense described in this first part of the dissertation. Since linguistic expressions and states of artifacts are intentional states, they should also count as representational states in my sense.

Finally, the analysis of intentionality in terms of representation has been the standard assumption in naturalistic accounts of intentionality (e.g. Dretske, 1986; Millikan, 1984; Papineau, 1987; Fodor,1990). Thus, adopting this terminology will provide us a more convenient way of talking and will facilitate the discussion with the long tradition of naturalistic accounts of intentionality.

For these reasons, I will assume throughout the thesis that ''X' is about X' means the same as ''X' represents X'. 'Representational state', hence, is intended as a semi-technical term for referring to intentional states.

1.1.3 *Semantic and Meta-semantic theories*

I have argued that the aim of this thesis is to provide a naturalistic account of intentionality but, if one carefully considers current and historical attempts of naturalizing intentionality, one discovers that there are two different questions a naturalistic theory of intentionality may attempt to address (Cummins, 1989, Sterelny, 1995, p. 254; Fodor, 2008, p. 217). On the one hand, we can seek to spell out why a certain representational state represents A rather than B; this question presupposes that a certain state is a representation and wonders what makes it the case that the content is *this* instead of *that*. Properly speaking, these accounts are *semantic* theories, since they seek to provide an explanation of why certain states refer to a certain entity, rather than to many alternatives. For instance, a semantic theory will assume that the linguistic expression 'Obama' is a representation, and will merely explain in virtue of what conditions 'Obama' refers to Obama rather than to Bush. Accordingly, these theories fall short of completely vindicating Local φ−physicalism, because they have to assume a semantic fact in the explanans; they claim that A means B iff A is a representation and condition X obtains. So they fail to explain why certain states mean something rather than nothing. While semantic theories might help to clarify certain aspects of representational phenomena, they only go halfway towards a complete naturalization of representation.

Alternatively, one might try to explain why certain states of affairs represent something at all, that is, one might try to find out what distinguishes representational states from non-representational ones. These are *metasemantic* theories, theories which purport to explain why certain entities (language, brain events, states of artifacts) possess semantic properties, while others (chairs, stones) lack them.[6] Metasemantic theories seek to account for the very existence of a mapping between representations and representata.[7]

This distinction is crucial for two main reasons. First, most naturalistic theories of content are semantic theories, but have often been mistakenly understood as metasemantic theories. Resemblance Theories, Causal Theories or Covariance Theories were usually not intended to provide an account of why some states represent something at all, but rather of what makes a certain representation be about A rather than B (e.g. see Fodor, 1987). As we will see, interpreting them as metasemantic theories threatens to render them trivially false, since most of them would have as a consequence that trees, clouds and rocks are representations. In that respect, I will show that some traditional objections against these theories fail to make this distinction.

6 Matthen (2006) also uses these expressions, although by 'semantic theory' he understands something slightly different. Similarly, what I call 'metasemantic theories' is labeled by Cummins (1989) 'theories of meaningfulness' and by Block (1997b) 'metaphysical semantic theories'.

7 Semantic and metasemantic explanations do not correspond to Dretske's (1988, ch. 2) explanations in terms of triggering and structuring causes. One could be confused at that point because Dretske also appeals to the difference between an explanation of why C causes M at all and explanation of why C causes M_1 rather than M_2. Both semantic and metasemantic theories look for structuring causes.

Secondly, in 2.2.4.1 I will argue that no satisfactory semantic theory can be offered, unless a metasemantic theory is also provided. Surprisingly, the (relatively easier) question posed by semantic theories cannot be answered without developping the (more difficult) metasemantic account. I will identify certain difficulties of extant semantic theories that can only be overcome by appealing to a metasemantic theory. Furthermore, I will argue that, in contrast to resemblance and causal theories, the fact that teleosemantics contains a metasemantic theory is one of the reasons it deals much better with some of the recalcitrant objections to other naturalistic approaches.

So much for the preliminaries. Let us move on to consider the main naturalistic approaches that have attempted to address in some way the problem of intentionality.

1.2 NATURALISTIC THEORIES

1.2.1 *Resemblance Theories*

If our goal is to naturalize representation and content, a first question we need to address is, What sort of physical relation can explain the fact that a state refers to another state? The first obvious candidate (historically and, perhaps, intuitively) is some kind of *resemblance* relation. In the same way a picture by Canaletto is about Venice because it resembles Venice (or so one might try to argue), one might defend that mental states and representations in general are about other states in virtue of some resemblance relation. Berkeley and Hume have traditionally been interpreted as holding a Resemblance Theory of (roughly) the following sort:

RESEMBLANCE A state R represents S iff R resembles S.

Obviously, in order for RESEMBLANCE to be clearly informative, one needs to spell out in more detail which kind of similarity must hold between R and S in order for R to represent S. Unfortunately, any intuitive way of cashing this relation out faces serious difficulties.

First, the notion of resemblance suggests that the representation and the representatum should be alike. But, of course, R and S cannot be *exactly* alike; when we perceive a cat, we do not literally have a cat in our heads. Rather, it was traditionally thought that the kind of resemblance between R and S is the same sort of resemblance relation that obtains between a picture and the objects depicted. Certainly, it intuitively seems that a picture of a cat resembles a cat to a certain degree. The problem, however, is that it seems so because the picture *looks to us* like a cat. Indeed, it is hard to find any property of the picture that resembles the cat other than the fact that it seems to be a cat to a subject (Cummins, 1989, p. 31-2; Prinz, 2002, ch.2). And, of course, we cannot appeal to this mental fact of subjects in order to define the relevant resemblance relation, since we would be defining a semantic property by mentioning a different semantic property, what would compromise our naturalistic goal.

A second option is to exclusively consider structural properties. One might try to define the resemblance between R and S by appealing to some structural properties shared by the two. However, we know that for any states R and S it is possible to find an infinite set of functions between parts of R and parts of S. For that proposal to work out,

we should find a way of picking out the relevant kind of structural similarity. Even if some gerrymandered mapping functions could be excluded, there seem to be always plenty of ways in which two structures resemble. So the existence of a structural similarity between representations and representata does not suffice for R to represent S.[8]

Summing up, a first problem with this approach is to find a satisfactory way of spelling out the kind of resemblance relation mentioned in Resemblance. It seems that the most obvious candidates fail (see also Pineda, 2012, p. 44).[9]

But the problem is not just that the resemblance relation is obscure. A simple example strongly suggests that resemblance (however we specify it) is neither necessary nor sufficient for representing.

On the one hand, the failure of sufficiency is quite straightforward. Suppose I take a picture of John and he happens to have a twin, Jonas. If John and Jonas are exactly alike, Resemblance predicts that the picture is of both John and Jonas. But that seems wrong, so resemblance is not sufficient for representing. More generally, any picture resembles a wide range of entities: a picture of a dog resembles a dog, but also a wolf, a cuddly toy, etc. Consequently, it is very unlikely that a resemblance relation suffices for representation (Prinz, 2002, p. 30).

A twist in this very same example shows that it fails to be necessary as well. Suppose John has plastic surgery done after I take a picture of him. In this case, the picture would resemble Jonas more than John, but we still want to say the picture is about John. Therefore, resemblance is not necessary for representing.

Finally, there are two more general kinds of arguments which have convinced most people that resemblance cannot play an important role in a theory of representation, but which I think are unsatisfactory (at least, as they are usually presented). First of all, some people argue that it is hard to think of any resemblance relation between the concepts VIRTUE, DEMOCRACY, TRUTH and their referents (Prinz, 2002, p. 30). So, the argument runs, resemblance theories will never be able to account for the content of these representations. The problem with this objection is that how to account for the representational properties of VIRTUE, DEMOCRACY is everyone's problem; for instance, it is not obvious how the things denoted by these concepts might covary or enter into causal or functional relations with our mental states. Similarly, if a solution is provided in terms of composition (see 6.5.3), the same proposal could be adopted by the supporter of Resemblance. So I do not think representations of abstract entities pose any special problem for resemblance theories.

The second failed objection is, I think, more interesting. The problem is supposed to be the following: Resemblance is a symmetric relation; if A resembles B, then B also resembles A. But, in general, if A represents B, then B does not represent A.[10] If representation were analyzable in

8 As we will see in 3.2.3, standard teleosemantic approaches also appeal to certain structural isomorphism between representation and representata. In that respect, they have the same problem as resemblance theories. The way they pick up the right mapping function is by resorting to evolutionary considerations. It is by developing a metasemantic theory (an account of that it is to represent something at all) that they manage to zero in on a particular mapping function.

9 One might try to deal with this problem by taking the resemblance relation as primitive, but seeking to naturalize intentional properties by appealing to a primitive and undefined notion of resemblance is very problematic. For one thing, the fact that A resembles B would be completely mysterious and hence it could hardly be considered a φ- fact.

10 In some very special cases, A can represent B and B represent A. For instance, 'this sentence has five words'.

terms of resemblance, the relation of representation would in general be symmetric. Since the relation of resemblance is asymmetric, representation cannot be reduced to resemblance (Goodman 1976, p.4; Prinz, 2002, p.31; Pineda, 2012, p. 45).

The flaw in this argument is that, as I suggested above, Resemblance is not intended as a metasemantic but as a semantic theory.[11] A pine tree resembles another pine tree, but someone endorsing Resemblance need not accept that one represents the other. Interpreting Resemblance as a metasemantic theory leads to a preposterous pansemanticism. Resemblance *assumes* a metasemantic theory (what makes a representation to be a representation), and hence it just takes for granted that R is a representation and S is not. As a result, the person who holds Resemblance and claims that my concept CAT represents cats in virtue of resembling them can easily deny that cats represent my concept CAT; she just needs to mention the fact that cats are not representations. This is one of the places where the distinction between semantic and metasemantic theories is essential.

Nevertheless, I think there is a satisfactory argument against Resemblance in the vicinity of the asymmetry argument. It is undeniable that some representations are about other representations, for instance, when someone stares at a picture or reads a book. In that case, since the representata are representations themselves, Resemblance predicts that each element represents each other. For instance, suppose one assumes Resemblance; then, if I am reading a book and my mental state JANE CALLED THE DOCTOR represents the sentence 'Jane called the doctor', then it resembles the sentence, and hence (by the symmetry of resemblance) the sentence should also represent my mental state. And that seems clearly wrong.

Finally, I would like to add a last argument against Resemblance that concerns cognitive science. Some cognitive psychologists working on the visual system have pointed out that abandoning the idea that mental representations have to resemble their objects was one of the main positive conceptual changes of modern cognitive science (Pylyshyn, 2003, 2007, Sternberg, 2009). For instance, psychologists had been puzzled for centuries by the fact that the projection of the visual field on the retina is inverted by the lens. It is well known that when light passes through the lens, the direction of the light is modified so that the image in the retina is, so to speak, upside down. Scientists were puzzled by the following question: How can we have an experience of standing objects if the projection of the image in our retina is inverted? For centuries psychologists unsuccessfully looked for the locus where the image is reverted, until they realized that a representation need not resemble its representatum (Sternberg, 2009).[12] As Pylyshyn (2007, p.2) writes:

11 Interestingly enough, there is some evidence that Hume distinguished the two questions. Garret (2008), for instance, argues that Hume probably held a sort of resemblance or causal (semantic) theory and a functionalist (metasemantic) theory.

12 Here is a quote from Kepler (quoted in Pylyshyn, 2007, p. 2) "I say that vision occurs when the image of the whole hemisphere of the world that is before the eye (...) is fixed in the reddish white concave surface of the retina. How the image or picture is composed by the visual spirits that reside in the retina and the [optic] nerve, and whether it is made to appear before the soul or the tribunal of the visual faculty by a spirit within the hollows of the brain, or whether the visual faculty, like a magistrate sent by the soul, goes forth from the administrative chamber of the brain into the optic nerve and the retina to meet this image, as though descending to a lower court—I leave to be disputed by [others]. For the armament of the opticians does not take them beyond this first opaque wall encountered within the eye."

> What made Kepler particularly pessimistic is that, despite years of trying, he could find no way within geometrical optics to deal with the problem of the inverted and mirror-reversed image on the retina. This puzzle left a generation of brilliant mathematicians and thinkers completely stymied. Why? What did they lack? It is arguable that they lacked the abstract concept of information, which did not fully come along until the twentieth century. The concept of information made it natural to see *right side up* and *upside down* as mere conventions, and allowed a certain barrier to be scaled because information requires only a consistent mapping and not the preservation of appearance.

Whether the notion of information is the key concept for solving the problem of representation will be discussed below, but I think it is pretty clear that the idea that mental representations need not resemble what they represent has been a powerful conceptual step forward in the evolution of science.

These arguments strongly suggest that the resemblance relation is probably not the crucial notion that will help us to naturalize intentionality.[13] Thus, I suggest to move forward to Causal Theories, which in contrast to Resemblance, still nowadays exhibit considerable support.

1.2.2 *Causal Theories*

The second sort of accounts that have been much influential in the recent literature are Causal Theories, which purport to analyze the phenomenon of representation taking *causation* or *nomological relation* as the key notion.

There are three main motivations for taking causality as the crucial relation. Historically, when the first recent naturalistic theories of intentionality arose, the causal theory of reference (Kripke, 1980) and the causal theory of perception (Grice, 1961) were popular views (see Dretske, 1981; Stampe, 1977). Thus, it was reasonable to think that if causation had yielded the right results in these fields, the same outcome could be obtained by using causation in the context of naturalistic accounts of intentionality.

Secondly, it seems to be empirically true that represented states usually cause (and covary with) their representations. That is, trees usually cause tokens of the concept TREE. At least, it seems trees cause more tokens of TREE than cows do. This intuitive nomological connection between representations and their referents is probably a key motivation for seriously considering some sort of causal relation as grounding a semantic relation.

13 It is worth mentioning that the revitalized interest on empiricist theories of concepts has not been accompained with a recovery of a resemblance (semantic) theories of intentionality. For instance, some people currently defend that humans think using the same vehicles that are activated in perception. Indeed, some have gone so far as to claim that we might think using images (Prinz, 2002; Raftopoulos, 2009a). However, even if some supporters of these imaginistic approaches to cognition maintain that the actual vehicles of thought are percepts (or perceptually derived states), they deny Resemblance and assume some version of the Causal Theory instead (see 5.2.2). So even if one of the main insights of traditional empiricism were right (that thinking is imaging), Resemblance still remains unattractive as a semantic (and, of course, metasemantic) theory.

Finally, Stampe (1977, p.82) puts forward a third motivation for considering causation as the key relation between representation and representatum:

> Our mental states are largely responsible for the success with which we occupy the world. That is evidence in favor of the idea our mental states represent the world accurately (to a large extent). However, that would be a miracle if we did not represent the objects in virtue of the fact that they cause our mental states.

In other words, the fact that we represent objects because they cause our mental states can easily explain why we happen to represent the world with great accuracy. This is an inference to the best explanation: if mental states represent certain things in virtue of being caused by them, that would provide a good explanation of (1) the fact that we often represent the world accurately (2) the fact that we act successfully. So causation seems to be a natural place to look.

In a nutshell, the Causal Theory of representation holds that it is in virtue of the fact that certain properties of an object are causally related to certain properties of a state that the latter refers to the former (Stampe,1977, p. 84).[14] The acceptance of this principle is a common feature of a large group of quite diverse theories. More precisely, there are three groups of theories that I classify under the label 'Causal Theories', which I think are motivated by roughly the same intuitions: the Crude Causal Account, Indication Theories and the Asymmetric Dependence Theory.

1.2.2.1 *Crude Causal Theory*

The Crude Causal Theory has probably not been defended in such a naïve form by anyone, but I think it is interesting to consider it because it suffers in its clearest form from the most important difficulties all theories of content must address.

The Crude Causal Theory claims that *being caused by* is a sufficient and necessary condition for representing:

Crude Causal Account R represents S iff R was caused by S.

Two caveats are important here. First, like resemblance, Crude Causal Account is a semantic theory, so it should be interpreted as specifying what should be the case in order for, say, CAT to refer to cats rather than to other things. Again, if Crude Causal Account were read as a metasemantic theory it would entail an unappealing panpsemanticism. If, for instance, it is granted that any physical entity has a cause, interpreting Crude Causal Account as a metasemantic theory would entail the preposterous view that any physical entity would be represent its cause.

The second important point is that there is an ambiguity in Crude Causal Account. There are two possible ways of understanding the criterion set up in the definition, which I will call an 'inclusive' and an 'exclusive' reading. On the inclusive interpretation, Crude Causal Account entails that if R is caused by S, then S *figures in* the content

14 Notice that if 'causally related' is read in a sufficiently broad way, one could classify certain kinds of teleological theories within this group. Nonetheless, I will follow standard treatments in interpreting more narrowly the notion of *causal relation* (see below).

of R, and (crucially) this is compatible with there being many other entities included in R's content. Thus, an inclusive reading of Crude Causal Account has it that if S and T cause R, then R has a single content, which includes S and T. In contrast, on the exclusive reading, the Crude Causal Account entails hat if R is caused by S, then S *is* the content of R. So, if both S and T cause R, then R has two different contents, *S* and *T*.

This distinction can be expressed more formally. The inclusive and exclusive interpretation of Crude Causal Account can be expressed in the following way (let us assume that we are allowed to quantify over contents, 'C' stands for 'causes' and 'F' for 'figuring in') :

Inclusive There is a content y such that, for any state x that causes a representation R, x figures in y. That is, $\exists y \forall x(Cxr \rightarrow Fxy)$

Exclusive For any state x that causes a representation R, there is a content y of R such that x figures in y. That is, $\forall x(Cxr \rightarrow \exists y Fxy)$

In short, the distinction between an inclusive and an exclusive reading is a distinction in the scope of the quantifier. Distinguishing between inclusive and exclusive readings is important because, as we will see, depending on the interpretation we take, Crude Causal Account gives rise to different difficulties.

As I said above, Crude Causal Account has obvious and important problems, and for this reason probably nobody has ever seriously endorsed it. However, I think it is worth discussing its drawbacks, because we will be able to formulate in its more clear form a set of problems that threaten all naturalistic accounts I will discuss in this and the next chapter.

1.2.2.2 *Error Problem*

Intentional states are states of affairs that can typically be accurate (or true) or inaccurate (or false) (Dretske, 1986, 1988, p. 64; Millikan, 2004). Sentences, thoughts or states of the thermometer are typically subject to cases of misrepresentation.[15] Misrepresentations usually occur when R is caused by something that is not in its extension[16], for instance, when a dog in a foggy day causes a token of CAT. Now, if Crude Causal Account is right and anything that causes a representation falls under its extension, then, *by definition* there is no S such that S causes R and S is not in the extension of R. If S does not cause R, then R does not represent S, and if S causes R, then S automatically falls under R's extension. Since it is not possible for R to be caused by S without R representing S, misrepresentation is impossible.

Let me illustrate the problem with the example of GOLD. Remember that we are looking for a theory that explains why GOLD means *gold* (rather than silver, bronze,..), that is, why gold falls under the extension of GOLD. Crude Causal Account claims that the extension of GOLD is determined by whatever causes tokenings of GOLD. But

15 However, the capacity of being inaccurate is not an essential property of representations, as many philosophers seem to suggest (e.g. Neander, 2012). For instance, arguably representations like 2+2=4 or *no surface is completely blue and completely red at the same time* cannot be false.

16 'Extension' is a technical concept. The extension of a representation R is the set of entities represented by R, i.e. the set of entities that *fall under* R, or the set of entities which R rightly *applies to*.

it is not difficult to see that not only gold causes tokenings of GOLD. In certain circumstances a piece of tungsten can also cause GOLD-tokens. Thus, if that happens, CRUDE CAUSAL ACCOUNT commits us to claiming that tungsten also falls under the extension of GOLD and, consequently, when we apply GOLD to tungsten, this is not a case of misrepresentation. But if this is not a case of misrepresentation, it is hard to see what else could be. The problem is that there is no logical space for error because candidates for error are transformed into non-errors by their mere occurrence.

More precisely, we can formulate the problem in the following way:

(Error Problem) A semantic theory suffers from the Error Problem if it does not allow for cases of misrepresentation

The problem is sometimes labeled 'the disjunction problem' (Fodor, 1990), because in this case claiming that misrepresentation is impossible amounts to saying that anything that causes a tokening of GOLD falls under its extension. That means that the extension of GOLD would consist of a long disjunction of things that actually cause tokenings of GOLD, the resultant meaning of GOLD being something like: *gold or tungsten or silver...* Later on we will have the chance of developing this argument in more detail. For the present purposes, it is enough if we see that this is a problematic consequence of the CRUDE CAUSAL ACCOUNT.[17]

1.2.2.3 *Adequacy problem*

Suppose a theory is able to solve the Error problem. As I formulated it, it will suffice if there are some cases where the representation is false. Suppose, for instance, that we have a principled reason for endorsing the claim that only natural kinds can be represented (see Ryder, 2006; Rupert, 2008). We would then formulate the new version of a Crude Causal Account in the following way: R represents S iff (1) S is a *natural kind* and (2) R was caused by S. Let us call such an account 'NATURAL KIND CRUDE', for *Natural Kind Crude Causal Theory.* Note that NATURAL KIND CRUDE can solve the Error problem (and also the 'disjunction problem', as formulated above) because there would be some cases where R would be caused by something that does not fall under R's extension. Assuming lemon chewing gums do not constitute a natural kind, when a lemon chewing gum causes GOLD that would qualify as an instance of misrepresentation.

However, notice that there would be many different natural kinds that would still cause tokens of GOLD. Silver, slate or bronze could still cause GOLD and, since they are natural kinds, they would fall under its extension. As a result, and assuming an inclusive reading of representing, GOLD would mean *gold or silver or bronze.* The problem

17 Two kind of confusions are usually found in many introductions to naturalistic accounts of content. On the one hand, this disjunction problem is usually considered a different objection from the Error problem, while I think it is a different way of expressing the same worry. On the other, the disjunction problem is usually classified as a particular version of the Indeterminacy problem, whereas I think it is a different sort of problem (see below 1.2.2.4). An instance of this confusion is the following passage by Burge (2010, p. 322):

> [The Disjunction Problem] is the problem of explaining conditions on representation that show that representation applies to one set of attributes rather than equally well to a set of alternative attributes.

here is not that error is precluded or that the content is disjunctive. The claim that GOLD means *gold or silver or bronze* is an unwelcome result because it clashes with our intuitions and scientific practice. So, even if Natural Kind Crude overcomes the Error problem, it still suffers from a serious drawback, which I call the 'adequacy problem'.

(Adequacy Problem) A theory suffers from the adequacy problem if the content it warrants greatly and systematically diverges from the content warranted by science and common sense.

The same idea was expressed by Neander (2006) when she said that a theory that entails that all mental states represent *Today is Tuesday* does not suffer from any problem of error or indeterminacy, but we would hardly claim it is satisfactory. The Adequacy Problem intends to capture this idea.

The Adequacy Problem is not supposed to imply that our theories can never contradict our intuitions. In order for a theory to be inadequate in the sense intended here, the disagreements between common sense and science on the one hand and the predictions of the theory on the other have to be *systematic.* Of course, there is some vagueness in the formulation of the principle, but nevertheless I think it is an intuitive criterion that might be useful for excluding theories that provide strongly counterintuitive results.

Notice that the objection is not just that the content is disjunctive (cfr. Rupert, 2008). Indeed, it is very plausible that many of our concepts have a disjunctive content. For instance, it is well known (at least, in philosophical circles) that JADE means *jadeite or nephrite*. Similarly, concepts like REPTILE or vague concepts are good candidates for having disjunctive contents. So the claim that a concept has a disjunctive content is not problematic *per se*. The difficulty of Natural Kind Crude (and Crude Causal Account) is that it attributes the wrong content.

Finally, let me mention that this objection derives from an inclusive reading of Crude Causal Account. In other words, Crude Causal Account suffers from the adequacy problem if we understand it as stating that any object that causes a given mental state is included in its (unique) content. If, instead, we interpret it as attributing different contents in accordance with different causes (that is, following what I labeled an 'exclusive reading') it suffers from the indeterminacy problem.

1.2.2.4 *Indeterminacy*

The last remarkable objection to Crude Causal Account is the so called 'Indeterminacy Problem', which was originally raised by Fodor (1990) against Teleological Theories of Content (see 2.3.3), but threatens any naturalistic account of content.

In a nutshell, the problem is that if we interpret Crude Causal Account exclusively, there are many different states of affairs R could be representing, and Crude Causal Account does not specify which one is the relevant content. For instance, my concept GOLD is caused by gold, but also by tungsten, silver,... similarly, it is also caused by waves impinging my retina, by gold-looking things... Since all these things cause R, an exclusive interpretation of Crude Causal Account entails that any mental state has many different contents.

More precisely:

(Indeterminacy Problem) A theory suffers from the indeterminacy problem if it warrants multiple content attributions in cases where science and common sense warrant a single content. (Martinez, 2010)

This problem will also be developed in some detail in 2.3.3.

Of course, there are several ways of trying to address the three problems of Crude Causal Account I presented. Some people think they can be solved by specifying which causal relations are content determining from those that are not. If a theory is able to specify *normal situations* in which the referent and (if possible) only the referent causes the representation,[18] we will move forward towards a solution. However, such an account would qualify as naturalist only if these *normal conditions* are specified in non-intentional terms. The challenge, then, is to provide a naturalistic account that is in a position to distinguish the good from the bad cases using non-intentional expressions. Several proposals in this direction have been offered. But, before presenting different theories that intend to address these issues, let me put forward a last (unsuccesful) objection to Crude Causal Account that is going to be specially important in the next chapter.

1.2.2.5 *Normativity*

The last difficulty of Crude Causal Account I would like to consider is that it might fail to account for the normative aspect of representations. In contrast to the three previous objections, I think Crude Causal Account can escape the normativity problem. Nevertheless, we will see that other theories, in particular Teleosemantics, should take it seriously. Let me explain.

Intentional properties are normative properties. Representations can be false (or unsatisfied), and when they are we have the intuition that something has gone *wrong*.[19] Misrepresentation is a special kind of failure. Now, Crude Causal Account purports to analyze the notion of representation by appealing to a causal relation between representations and representata, but it is hard to see how this causal relation can ground the normativity involved in representational states. Crude Causal Account might not be able to explain the normative difference between on the one hand, being accurate and inaccurate, and on the other, being heavy or light. The former seems to involve a normative aspect that is missing in the later.

Let me phrase the worry in a different way. Suppose Crude Causal Account is true and the fact that cats cause tokenings of CAT is sufficient and necessary for CAT to represent cats. Imagine that Crude Causal Account can solve the Error problem, and hence can distinguish cases of misrepresentation from the rest. Still, one might ask, Why is having a misrepresentation *wrong*? Why (all things being equal) should we avoid producing false representations? Intuitively, causal relations alone fall short of accounting for this normative aspect of intentionality.

18 If we specify a set of normal conditions where the referent causes the representation but there are also a bunch of other states that also cause it, then we might be able to overcome the Error problem, because abnormal conditions would allow for misrepresentations, but we will probably not solve the Indeterminacy problem or the Adequacy problem, since the representation will still have too many contents and have an inadequate content.

19 In a sense, the mere fact that states have accuracy conditions is a normative fact. That is, the Error problem is also a failure to account for a certain normative aspect of representations. However, here I am pointing at a different normative aspect of representations.

Now, I think there is a good reply available to anyone endorsing Crude Causal Account, namely that this approach is intended as a semantic theory, not as a metasemantic one, and the problem of normativity has only to be addressed by a metasemantic theory. Like other semantic theories, Crude Causal Account assumes that we have an independent theory of what representations are; and, since representational states are normative, Crude Causal Account takes for granted we have an account of what endowes representational states with this normative import. Thus, Crude Causal Account can happily grant that we have an independent story to tell about normative properties. Therefore, the supporters of this view can just reply that it is not their job to address the normativity problem.

Of course, rather than solving the objection, this reply just passes the bug to metasemantic theories. Thus, the normativity objection does not affect the rest of accounts I am going to consider in this chapter, but it is an important challenge for Teleosemantics, which is intended as a semantic and metasemantic theory. So there is a further problem any satisfactory *metasemantic* account of intentionality must be able to address:

(Normativity Problem) A metasemantic theory suffers from the normativity problem if it cannot account for the normative difference between cases of successful representation and cases of misrepresentation.

I will discuss in the next chapter whether teleosemantics can overcome this problem.

I think these are the four most general objections faced by naturalistic accounts of content and representation. Let us consider now some other proposals and assess whether they satisfactorily address these issues.

1.2.3 *Indication Theories*

I outlined some general motivations for causal theories, which also lend support to indication theories. I would like to add two additional motivations for indication theories: artifacts and cognitive science.

We are familiar with many devices like barometers, which (among other things) represent cloudy weather. However, barometers do not represent cloudy weather in virtue of the fact that the latter causes the first, but in virtue of a correlation between the two. Usually, when the barometer is in a certain state, cloudy weather ensues. In that case, of course, the covariation between the barometer state and clouds is underpinned by a common cause: low pressure. But one might argue that the representational relation is based on the strong correlation between the two states. These and similar cases suggest that representation might be grounded on covariation, rather than direct causation.

A second more recent motivation comes from cognitive psychology. When neuroscientists (among others) investigate what a given neuronal structure represents (for instance, in single-cell recording), they usually try to establish what kind of stimulus most prominently covaries with neuronal activation (see 4.1.4.2). So, prima facie, it seems that cognitive psychology assumes that a brain structure represents whatever stimulus it covaries with (Eliasmith, 2000, 2003; Jacob and Jeannerod, 2003).

Accordingly, Indication theories claim that representing should be analyzed in terms of covariation, where a state indicates another iff the

first covaries with the second to a certain degree. In this case, different notions of co-variance imply different theories of representation and lead to different problems. Let us consider some of these proposals.

1.2.3.1 *Strong Indication*

In his seminal work on naturalistic theories of representation, Dretske (1981)[20] famously endorsed the claim that a state R represents another state S if the probability of the second given the first is 1 (assuming that certain background conditions hold). More precisely:

STRONG INDICATION Structure R has the fact that t is F as its semantic content iff R carries the information that t is F in digital form (Dretske, 1981, p.177)

Where (1) R carries the information that t is F in digital form iff it is the most specific piece of information that it carries about t and (2) R carries the information that t is F iff P(t being F | R) = 1.

It is a commonplace in the literature that this theory suffers form the same Error problem that affects Causal Theories in general. If R represents that S (let 'S' refer to the state of affairs constituted by t being F) only when the probability of S obtaining given R is 1, it will never be the case that R represents that S and S does not obtain. And since this is precisely the situation in which R is false, STRONG INDICATION does not allow for misrepresentation. Nonetheless, notice that it fails to account for misrepresentation for a different reason from CRUDE CAUSAL ACCOUNT. While CRUDE CAUSAL ACCOUNT included any cause into the content (which also lead to the adequacy and indeterminacy problems), STRONG INDICATION restricts very much the kind of states that can be represented. So, prima facie, it is not obvious that it also falls prey to problems of adequacy and indeterminacy (but see below).

Dretske was aware of this problem, and tried to solve it by distinguishing a learning period from a post-learning period, such that only the first is content-determining.[21] The idea is that during the learning period the probability of the represented state of affairs obtaining, given that the representation obtains, is 1; however, after a representation is learned, this probability might be reduced, and that is what allows for misrepresentation. Once a subject learns that R refers to S, he can make mistakes. The learning period is the reference-fixing situation, while the post-learning period leaves room for wild causes. So, if a learning period can be distinguished from a post-learning period when recruiting representations, it seems STRONG INDICATION will be able to account for misrepresentation (at least, in the post-learning period).

However, the learning period solution has two serious problems. First, such an account requires a sharp distinction between a period where the probability of S given R is 1 and another period where it is less than 1, and it is empirically implausible that in general such a period exists in representational systems. As a matter of fact, representational systems lack an infallible learning process that is clearly

20 It is common to classify Dretske's naturalistic theory which appears in *Explaining Behavior* (1988) and *Naturalizing the Mind* (1995) as an Indication theory, since he uses the notion of indication. Nonetheless, I will classify it as a particular version of Teleosemantics. The reason is that in these works he strongly relies on the notion of function (roughly, 'X' represents X if 'X' has the function of indicating X).

21 Dretske also uses learning in his metasemantic theory. In particular, he employs it in order to distinguish mental representation from mere low-level functioning sensitivity (e.g., see Dretske, 1988). For a convincing criticism, see Burge (2010, p. 305-307).

differentiated from a period of mistakes. For instance, that seems to be true of any organism learning representations of food or danger (see 2.2.3).

Secondly, Dretske does not explain how representational systems manage to be infallible during certain period of time. At some point, he suggests that a faultless correlation between R and S can be guaranteed by introducing a second device (a *teacher*) which rectifies any miscorrelation. So, in order to ensure an infallible learning period, a second device needs to be introduced that intentionally corrects the mistakes of the learning representational system. Perhaps such a framework can to ensure an infallible connection between R and S, but a proposal that explains semantic properties of representations by appealing to further intentional facts (the supervisor) is clearly unsatisfactory from a naturalistic point of view. The φ- facts appealed to in the *explanans* would include a semantic fact.[22]

1.2.3.2 *Weak Indication*

Rather than postulating an infallible learning period, one can solve the Error Problem by simply lowering the probability that is required for R to represent S. Notice that (contrary to standard expositions of indication theories; e.g. Aizawa, 2010) once the probability of the representatum obtaining given the representation is below 1, the Error Problem is solved. To accommodate misrepresentation, it suffices if a theory is able to distinguish cases where the representation is accurate from cases where the representation is not accurate. If we assume that, say, a state R represents a state S iff $P(S \mid R) = 0.8$, there is room for misrepresentation. R will misrepresent S when R obtains and S does not, which happens a fifth of the times R is tokened. So, in contrast to Crude Causal Account and Strong Indication, if the probability that is required for R to represent S is below 1, then the Error Problem is avoided.

But there is another important aspect that needs to be taken into account when specifying the probability that needs to be included in a naturalistic theory of content. All theories I have discussed so far are semantic theories, so they purport to establish a general criterion for any state to represent A (rather than B). However, in the natural world there is an impressive amount of diverse representational systems, which exhibit different degrees of covariation between the representation and its representata. The correlation between signals and what they signify can be extremely varied. Think, for instance, about warning signals. Alarm calls between animals increase the probability of a predator being around, even if most of the time there is no threat around (see 3.3.3). So if all these signals are to count as representations (and remember we are looking for a general account of intentional states), then the threshold of covariation has to be very low. Raising the probability will leave out certain states that we want to count as full-blown representations. Hence, a plausible indication theory, which is able to deal with the Error Problem and account for all kinds of representation might be along the following lines:

Weak Indication A state R represents state S iff $P(S \mid R) > P(S)$

22 One way of solving this problem is by identifying a non-intentional corrector, but it is hard to think of any non-intentional supervisor that can ensure an infallible learning period (surely, evolution cannot be such a mechanism).

In other words, R represents S iff the probability of S obtaining given R is higher than the probability of S obtaining, i.e. the fact that R obtains increases the probability of S occurring. I will call this sort of correlation 'weak covariance', in order to distinguish it from the stronger covariance contained in STRONG INDICATION.

As I said earlier, this approach solves the Error Problem; since the probability of S obtaining given R is less than 1, there will be some cases where R is tokened, R means S but S is absent. However, solving the Error Problem and being able to account for all kinds of representations comes at a price.

First, we saw that in order to allow for misrepresentation, the required probability has to be below 1. Furthermore, since the semantic theory we are looking after has to determine the necessary conditions for *any* representation to be about a certain state (rather than about another state), the probability required has to be very low, because there are many different kinds of representational systems in the natural world, which have a very low degree of covariation with their referents. The problem is that both desiderata push us towards a very low threshold for R to represent S, and once we move in that direction the adequacy and indeterminacy problems show up. If for R to represent S we only require that tokens of the first increase the probability of the second obtaining, R will be representing many different states of affairs. In other words, STRONG INDICATION is too strong and hence it suffers from the Error Problem, but WEAK INDICATION is so weak that it suffers from the adequacy and the indeterminacy problems. For any representation R, there are plenty of states that can play the role of S in WEAK INDICATION. So it is hard to see how covariation can provide a satisfactory approach that be able to overcome these worries.

1.2.3.3 *Relative Indication*

Rupert (2008) and Eliasmith (2000) have recently defended an indication theory which differs from STRONG and WEAK INDICATION in that it brings forward a comparative dimension among candidates for representata (it is worth reminding that we are looking for a semantic theory, that is, an account that be able to explain why A refers to B rather than C). Instead of specifying an absolute degree of covariance that must hold between the representation and the representatum, they establish a comparative criterion; A represents B rather than C if (roughly) B activates A more strongly than C activates A. In Rupert's (2008, p. 362) words:

> RELATIVE INDICATION R has as its extension the members of natural kind[23] Q if and only if members of Q are more efficient in their causing of R than are members of any other natural kind.

The details of how to specify efficiency might vary among authors, but the basic idea is the same. Rupert, for instance, provides a precise definition of *efficacy*:

> For each natural kind property Q_i calculate its PRF [past relative frequency] relative to R; divide the number of times

23 Rupert restricts his account to natural kinds. An interesting question is whether Rupert can explain how organisms can represent things other than natural kinds, without introducing intentional notions. Since I think the theory faces more serious objections, I will not press this point any further.

> an instantiation of Q_i has caused S to token R by the number of times an instantiation of Q_i has caused S to token any mental representation whatsoever. Then make an ordinal comparison of all $Q_j$relative to that particular R; R's content is the Q_j with the highest PRF relative to R". (Rupert, 2008, p.362)

Similarly, Eliasmith (2000, ch 4) appeals to the notion of *highest statistical dependency*, which he also analyses in terms of frequency.

By introducing this comparative aspect among stimuli, these accounts are able to avoid the problems of Strong and Weak Indication. On the one hand, Rupert's view can account for misrepresentation, since the natural kind that is more efficient in causing R can fail to cause it in a given occasion. Even more: this view is compatible with the representation being false most of the time, since the stimulus that has the highest statistical frequency may be missing very often (Eliasmith, 2000, p. 34). Secondly, in each occasion it seems to be able to provide a quite definite content: R represents the natural kind that most efficiently causes it. Rupert avoids having to draw a general threshold (which, as I argued, has to be very low) by putting forward a comparative threshold among states. So, prima facie, it seems to overcome all the objections I have considered so far.

Unfortunately, this proposal still faces a set of important problems. First, it seems that in many cases it yields the wrong content attributions. For one thing, some neurons in the visual cortex may be the most efficient cause of a representation or concept CAT. More generally, among all the events that form a causal chain from the external object to the representational states, Relative Indication does not tell us which is the right level of distality (or, rather, it takes the most proximal state as the one being represented).

Secondly, assuming that representational content is determined by whatever activates more strongly (say) a given brain state leads to a different kind of counterintuitive results. A clear counterexample is presented by Martinez (2010), who points at the existence of supernormal stimuli. Supernormal stimuli are stimuli that produce in an animal a response that is stronger than would be evoked by the natural stimulus it resembles (Tinbergen, 1960; Sterelny, 1995, p. 257). For instance, in some birds the incubation behavior is stimulated by the presence of an egg; the larger the egg the larger the stimulus. Since the brain structure responsible for the incubation behavior responds more strongly to the presence of a supernormal stimulus like (say) an ostrich egg, many birds would count as representing the presence of an ostrich egg. That result is surely wrong.[24]

Finally, it seems that Rupert is taking the wrong direction of explanation. Certainly, we can expect that in general the representation R is mostly activated when confronted with its referent S. In other words, it is likely that we usually find a correlation between satisfying Relative indication, and the fact that R represents S. However, it seems plausible that the fact that R represents S *explains* that Relative indication holds and not vice versa. Rupert is right in that it is reasonable to attribute content S to R when R reacts more strongly to S than to any other thing. But this is far from accepting that Relative indication

24 This objection could be overcome by considering only those stimuli that were present around the organism during the evolutionary past of its species. A similar proposal will be discussed in 4.1.4.2.

grounds the fact that R represents S (This issue will be discussed in 4.1.4.2).

A different way of putting the same worry is by pressing on the justification for the right-hand side of RELATIVE INDICATION: what explains that R is more strongly activated when confronted with S than with T? At some point, Rupert seems to be suggesting that children get corrected by other people, in a way that R slowly gets more and more correlated with S and less with other things. That might be an improvement over Dretske's own account of learning, since it avoids having to draw a clear-cut line between a learning and a post-learning period, but it still faces the naturalistic worry; what explains the correlation is that someone intentionally directs the subject onto its referent. That is unacceptable from a naturalistic point of view.

In conclusion, I think it is very unlikely that something like RELATIVE INDICATION provides a satisfactory account of representation. Nevertheless, despite the fact that I think RELATIVE INDICATION fails as a semantic theory, let me advance that it will play an important role in the second part of the dissertation. The reason is that I think that one of its motivations (the fact that neuroscience seems to follow RELATIVE INDICATION) has to be taken seriously. In 4.1.4.2 I will argue that an account along the lines of RELATIVE INDICATION can be used as a methodological principle in order to identify the representational contents of brain states. So, whereas I will think that something similar to RELATIVE INDICATION provides a useful methodological strategy for addressing complex cognitive systems, it falls short of providing a semantic (or metasemantic) theory of representational content.

Summing up, while I think that there is an important intuition in favor of indication theories that I will try to rescue in following chapters, I think covariation cannot be the key relation that makes the naturalization of content possible. We need to look for something else.

1.2.4 *Asymmetric Dependence Theory*

The third causal account of content I will consider is the Asymmetric Dependence Theory, which was originally put forward by Fodor and has been accepted by few philosophers (Margolis, 1998; Stalnaker, 1984). The basic idea of the Asymmetric Dependence Theory is that there is a dependence relation of the causes that do not determine content on the causes that do. More precisely, if GOLD means gold, this is because the following dependence relation holds: non-gold causes GOLD tokenings because gold causes GOLD tokenings. According to Fodor, this is true in virtue of the fact that certain counterfactual relations hold: If gold did not cause GOLD, non-gold would not either, but if non-gold did not cause GOLD, gold would still cause it.

Again, notice that the Asymmetric Dependence Theory is not a metasemantic but a semantic theory. In particular, Fodor accompanied the Asymmetric Dependence Theory with a functionalist account of propositional attitudes. According to him, a state is a belief if it plays a certain role in a subject's economy, but it is a belief about A (and not about B) because A satisfies ASYMMETRY (Fodor, 1987).[25]

25 Millikan (1993 p.84) argued that it seems to follow from Fodor's view that a state can be a belief without being a belief of any particular thing, if it satisfies the functional role but it does not satisfy ASYMMETRIC. This is a sensible worry that could be extended to all accounts that detach a semantic theory from a metasemantic one.

More formally, according to the Asymmetric Dependence Theory:

Asymmetry

R represents S iff:

1. *S cause Rs* is a law.
2. For all Ts that are not Ss, if Ts actually cause Rs, then the Ts causing Rs is asymmetrically dependent on the Ss causing Rs.

Where *Ts causing Rs* is asymmetrically dependent on Ss causing Rs iff breaking the law that links Ss to Rs breaks the law that links Ts to Rs but not vice versa.

Interestingly enough and in contrast to other naturalistic theories, no philosopher besides Fodor has actually tried to develop a version of the asymmetric dependence theory, even if some have claimed it is a promising starting point (Margolis, 1998; for a summary of criticisms, Adams and Aizawa, 2010). Consequently, in this section I will exclusively be concerned with the Fodorian account.

On the other hand, it is worth mentioning that, pace Fodor (1990, p. 120), information does not play any significant role in his theory. Let me present a case that makes that vivid. Fodor wants to attribute content to representations of uninstantiated properties, such as PHLOGISTON, and *prima facie,* it might seem he has a way of doing that, since there can be an asymmetric dependence between laws that involve uninstantiated properties. For instance, the law that links oxygen to PHLOGISTON, might be asymmetrically dependent on the law that links phlogiston to PHLOGISTON, because in the most proximal world where there is phlogiston but not oxygen, the former still causes PHLOGISTON, but in the most proximal world where there is oxygen but not phlogiston, oxygen does not cause PHLOGISTON.[26] Now, as it is usually understood, a state A carries information about state B iff A has been caused by B (Dretske, 1981) or A covaries to a certain degree with B (Millikan, 2004; Skyrms, 2010, p. 37). Since PHLOGISTON means *phlogiston* in virtue of an asymmetric dependence theory, but phlogiston neither causes nor covaries with PHLOGISTON in the actual world, there is no sense in which the meaning of PHLOGISTON derives from any information carried by this concept. More generally, the fact that there is an asymmetry between laws does not straightforwardly imply that there is an informational relation between causes and representations. Hence, I think Fodor's Asymmetric view should not be considered an informational theory.[27]

There are two crucial advantages of the Asymmetric Dependence Theory over previous accounts. First, since it only appeals to dependence relations between laws, it has a way of distinguishing content determining causes from *wild* causes. When a cow causes a tokening of

26 Of course, I am not assuming that this is a satisfactory answer. For one thing, it is not clear that the last claim is true. My goal here is just to point out that, in contrast to standard classifications, Fodor's asymmetric dependence theory should not be classified as 'informational'.

27 It could be argued that the theory is still informational because PHLOGISTON would have carried information about phlogiston in those nearby possible worlds where phlogiston exists. However, notice that, on this interpretation, PHLOGISTON in the actual world means phlogiston in virtue of an informational relation holding in other possible worlds. Thus, even on this reading, it is granted that there is no informational connection in the actual world between phlogiston and PHLOGISTON. This is a substantive difference between Fodor's account and classical causal-informational accounts.

HORSE, HORSE is misrepresenting because the law that links cows to HORSE asymmetrically depends on the law that links horses to HORSE. So it clearly solves the Error Problem. Secondly, Fodor thinks that it also solves the Indeterminacy Problem (Fodor, 1990, 105). Let me illustrate this claim with a classical example. Frogs are endowed with a prey-catching- mechanism that reacts to flying preys by detecting black moving things.[28] Fodor view seems to entail that the content of the states produced by this mechanism is *there is a little black thing*. Little black things would still cause the firing of the frog's neurons even if flies were not their cause.[29] Finally, it achieves these results by specifying a dependence relation that can be spelled out in non-intentional terms. So his account seems to have the right naturalistic credentials.

All this make Asymmetry very compelling. However, several reasons suggest that it is probably unsatisfactory.

objections The first problem faced by Asymmetry concerns concepts like JADE. As we know, jade is actually composed of two different substances, namely jadeite and nephrite. Let us assume that jadeite and nephrite look exactly the same, so that they can only be distinguished using scientific methods. Now, according to asymmetry the question that is relevant for determining content is the following: if jadeite did not cause JADE tokens, would nephrite cause JADE? The answer seems to be affirmative. If by a freak atmospheric change pieces of jadeite alter its appearance so that they do not cause tokenings of JADE any more, nephrite would still cause them.[30] But, similarly, if nephrite did not cause JADE tokens, jadeite would still cause tokenings of JADE for exactly the same reason. The upshot is that neither the law that links jadeite and JADE is asymmetrically dependent on the law that links nephrite and JADE, nor is the second asymmetrically dependent on the first. So JADE neither means jadeite or nephrite, which is certainly false.

Fodor could possibly reply that we are considering here the wrong kind of causes. He could then argue that we should evaluate the counterfactuals by considering *being jade* (that is, *being jadeite-or-nephrite*). Certainly, it seems that non-jade causes tokenings of JADE only because jade does, so the asymmetric dependence would hold and Asymmetry would have it right. However, we might wonder why should we specify the counterfactuals using the property of *being jade* instead of the properties *being jadeite* and *being nephrite*? After all, jadeite and nephrite are natural kinds, and it is usually assumed that ideally scientific laws range over natural kinds; so condition (1) of Asymmetry would most naturally be satisfied by these two substances rather than by jade, which is not a natural kind. Fodor could reply that we evaluate the counterfactuals using *being jade* (*being jadeite-or-nephrite*) and not using *being jadeite* and *being nephrite* because we know that JADE means jade. However, we would insist this is precisely what we are trying to settle; the representation's meaning cannot be merely assumed in the explanans. So I think the right way of applying Asymmetry is between

28 For more on this example, see 2.3.3 and 4.2.2.

29 One might argue that this claim is false, because if black dots had not been flies, frogs would not have evolved a mechanism for detecting them. In order to avoid this sort of counterexamples, Adams and Aizawa (2010) suggest to add a further condition in Asymmetry: that the dependence is synchronic. I will not go into these details.

30 Notice that the synchrony restriction is also relevant here. If we did not have this constraint, we should consider cases where jadeite had never caused JADE, which are far more difficult to evaluate.

jadeite and nephrite and, if so, then Asymmetry entails that JADE has no meaning.[31]

Leaving ambiguous terms apart, I think the most serious problem for Asymmetry is that it warrants extremely counterintuitive contents to representations. Here the problem is not indeterminacy (the content is highly determinate) but adequacy. There are two ways of developing this objection.

First, the law that links horse and HORSE is asymmetrically dependent on the law that links horse-looking things and HORSE, since breaking the first connection does not imply that we break the second, while breaking the second would surely break the first. So, it seems that, according to Asymmetry, HORSE means *horse-looking thing* rather than *horse.* Similarly, consider the kind of light impinging my retina that produces a sensation as if there were a horse in front of me; call it 'L'. The law that links horse-looking things and HORSE is asymmetrically dependent on the law that links L and HORSE, since breaking the latter would entail breaking the former, but not vice versa. This argument could be iterated so that Asymmetry seems to warrant a representation of the most proximal cause. We would never be able to represent distal things (see Fodor, 1990, p.110).

Indeed, suppose we have a way of fixing this problem, perhaps by appealing to different situations where we token the same representation but have different proximal inputs (Fodor, 2008; see also 4.2.3.3). A parallel problem remains at the different levels of abstractness. Imagine we have a way of ruling out light impinging my retina and horse-looking things when specifying the content of HORSE. Still, it is the case that in nearby possible worlds in which animals do not cause HORSE, horses do not cause HORSE, and in nearby possible worlds in which horses do not cause HORSE, animals still do (probably, some animals that closely resemble horses). So the law between horse and HORSE is asymmetrically dependent on the law between animals and horses. Similarly, the law between animals and HORSE is asymmetrically dependent on the law between living beings and HORSE, and so on. The upshot is that if we took Asymmetry to its last consequences, any representational state would be representing the most general category- *there is something*.

Finally, I think that Asymmetry suffers from the same problem as Relative indication, namely that it reverses the order of explanation. Certainly, it seems that in general we can expect to find very often an asymmetric relation between the the law that links representatum and representation and the law that links other causes and the representation. But what *explains* this asymmetric relation is the fact that the representation has such and such content.[32] Here is a different way

31 David Pineda has suggested to me an interesting reply on behalf of the defender of the Asymmetric Dependence. One might argue that this problem could be avoided by reformulating Asymmetry in the following way: R represents S iff: (1) *Ss cause Rs* is a law (2) there is no T such that T causes R and Ss causing Rs is asymmetrically dependent on Ts causing Rs. One might argue that, since only nephrite and jadeite are not asymmetrically dependent on any other substance, this modified approach seems to get the right result. However, we will see below that nephrite and jadeite are indeed asymmetrically dependent on many other properties.

32 Cummins (p. 1989 p. 59) was probably pointing at a similar worry, when he wrote:

> The theory of asymmetrical dependence inverts the explanatory order: MOUSEs are wild when caused by shrews not because the more basic causal connection is with mice, but because MOUSE expresses the property of being a mouse -something that might well do even if the dependence were symmetrical.

of putting the same idea: what explains that there is an asymmetric dependence relation between the law that links dogs to CAT and the law that links cats to CAT? Fodor has never offered such an explanation. I can think of only one satisfactory explanation: the fact that CAT means cat (this objection will be extended in 5.2.2).[33]

1.3 CONCLUSION

In the first part of this chapter I clarified the main goal of the dissertation: trying to provide a naturalistic account of semantic facts. In particular, in this thesis I will argue in favor of LOCAL φ−PHYSICALISM by showing how (a privileged set of) semantic facts metaphysically supervene upon (and, indeed, reduce to) φ−facts. Developing these ideas in detail will take us the rest of this dissertation.

In the second half of the chapter, I have surveyed some candidates for providing a naturalistic account of representation and content, which have revealed a set of problems that any theory should address. The fact that Resemblance and Causal Theories fail to adequately deal with these difficulties (among others), show them to be unsatisfactory as semantic theories of content. Nevertheless, I think some of the ideas defended by these accounts are compelling and indeed I am going to use some of them in the following chapter. In particular, the idea of structural resemblance, the notion of RELATIVE INDICATION and the appeal to causal relations will be employed in different ways in the following chapters.

There are now two tasks at hand. First we need to find a satisfactory semantic theory that be able to overcome the Error, the Adequacy and the Indeterminacy Problems. But remember that this is only the first part of the project. We still need a metasemantic theory, which be able to explain what representations are and solve the Normativity problem. Otherwise, the naturalistic project would be incomplete. I think a teleological theory is the most promising approach that can fulfill all these desiderata. The rest of Part I is precisely devoted to describe a satisfactory teleological theory of representation.

33 As a final remark on Fodor, let me mention that sometimes Fodor seems to be claiming that his theory is a metasemantic account of representation (not only a semantic one), and for this reason he has sometimes addressed the problem of Normativity. In particular, in Fodor (1990) he wants to introduce the notion of (biological) function in ASYMMETRY in order to account for this normative dimension of representations. However, it is not easy to see how the notion of function can fit into his framework. Furthermore, if, after all, ASYMMETRY needs the notion of function to account for representation, the most natural way to go is to adopt a teleological view, which proceeds without asymmetric relation. If, as I will argue later, the notion of function (and sender-receiver structure) suffices for explaining representation, ASYMMETRY would then be unnecessary.

2

TELEOSEMANTICS

The main goal of this second chapter is to describe the Teleosemantic Theory and to show that it can solve the problems of previous accounts. I will set up the main ideas of the framework that I will be using in the rest of the dissertation.

The distinctive aspect of Teleosemantic or Teleological Theories is their use of the notion of *function* (Millikan, 2004, chapter 5-6). There are two key aspects of the concept of function that, prima facie, seem promising for any naturalistic theory of content. First of all, the notion of function might help us to explain how misrepresentation is possible and hence to solve the Error Problem; the strategy is to reduce misrepresentation to malfunction and then to explain away the notion of function in non-intentional terms. If that can be done, that would constitute a considerable step towards the naturalization of content (Millikan, 2004, p. 63). Secondly, this notion might help us to solve the problems of adequacy and indeterminacy. Perhaps a suitable notion of function will allow us to say that while there is an infinite set of objects that cause a certain representation, the state represents only some of them because its function is to covary with one and not the other (Papineau, 1993, ch. 4). Finally, the concept of function is essentially normative (traits *are supposed* to perform their function), so it looks like a promising place to look for overcoming the Normative problem. These intuitions, which of course are quite imprecise and sketchy, constitute the original motivations for teleological theories. This chapter is intended to show that these intuitions are approximately true, even though things are more complicated than they might seem at first glance.

The chapter has three main parts. First of all, I present the debate on the notion of function, and define in more detail the relevant concept of function that will be employed in the rest of the dissertation. In this first part I also define three central notions of the teleosemantic account: 'Reproductively Established Family', 'Selection for' and 'Darwinian Population'. In the second part of the chapter, I show how this notion of function can be used in order to provide a naturalistic account of content. I will provide the first definition of teleosemantics, which will be refined in several ways in the following chapters. Finally, in the last part I argue that the teleosemantic account I put forward can deal with the four problems set up in chapter 1.

2.1 FUNCTION

The first thing we need in order to develop a teleological account is to define a satisfactory notion of function. Unfortunately, function itself is a very controversial concept (Godfrey-Smith, 1993). So, before any teleological theory can be defended, we need to define and argue for a particular interpretation of this notion. This is what I intend to do in this first section.

2.1.1 *The Project*

There are at least three different ways of approaching an analysis of the notion of function. So, before presenting and discussing certain theories, we should be clear about what the goal of the debate is.

NOMINAL DEFINITION The first project is to engage in conceptual analysis of the expression 'function', as it is used in different scientific domains, specially in the biological sciences. On this approach, a theory of function searches for the criteria of application that people generally have in mind when they use the term 'function' (Neander, 1991). Roughly, the main goal is to unravel the set of necessary and sufficient conditions that scientists *think* an entity must satisfy in order for them to ascribe it a certain property. For instance, if scientists think that a device D has a property F iff D does E, then according to the nominal definition of F, a device D has F iff D does E.

REAL DEFINITION A second project, which differs from the first one but it has usually been confused with it, assumes that *having a function F* is a natural property like *being water* or *being a kangaroo* (Millikan, 1989). Accordingly, the goal of this second kind of theories of function is not to discover what scientists have in mind when they attribute functions, but rather to spell out in virtue of what features certain entities instantiate the property *having a function F.*

Obviously, the most natural way of addressing this question is by looking at science, so in many cases the project of finding a Nominal Definition and the project of discovering a Real Definition usually coincide in their results. Nevertheless, it is important to notice that, in some cases, both projects will give us different results; it might well be that scientists attribute functions by appealing to certain properties but what really grounds the fact that this entity has a function is something else.

There are two ways a Nominal Definition can differ from the Real Definition. First of all, scientists might be wrong about the defining properties of F; in that case, the Nominal Definition would describe what scientists have in mind when attributing F and the Real Definition would describe what really defines F. But, more interestingly, both could be right at the same time and define F in different ways. That can happen, for instance, if scientists attribute F by relying on a symptom; the Nominal Definition would then appeal to a property that is used by scientists in order to attribute F and the real definition would appeal to the real property that makes a certain entity have F. This last situation can be illustrated with an example: a theory of measles might try to specify how doctors attribute measles to patients (by seeing red patches onto their skin) or might try to specify the conditions in virtue of which someone has measles (having *paramyxovirus*).

Notice that the project of discovering Real Definitions might be complicated by the fact that some people think that 'function' refers to two different properties (Godfrey-Smith, 1993) in the same way 'jade' does. Nevertheless, the goal of the project is still clear: to specify the conditions that must obtain for an entity to instantiate the property or properties that scientists pick up when they use the notion 'function'.

STIPULATIVE DEFINITION The last project that one might seek to carry out when looking for a theory of function is to introduce a technical notion 'function', which is supposed to do a certain job within a theory. In this case, a stipulative definition is provided, which is not supposed to correspond to how scientists attribute functions or to any natural property.

Given these different way of addressing the question of functions, are there any reasons for teleosemanticists to opt for one or another analysis? Let me argue why teleosemantics should look for a real definition of function.[1]

2.1.1.1 *Teleosemantics and Functional Analysis*

There are two important reasons for assuming that teleosemantics must be concerned with the second kind of definition of 'function'.

The first obvious reason is that the goal of teleosemantics is a metaphysical naturalization of a certain phenomenon (as defined in 1.1.1), so we are primarily interested on what there is, rather than on any properties that scientists think there is. In this thesis, we will pick up a set of properties and will show how they supervene on other properties. In particular, we will use the notion of function in order to identify certain structures that give rise to semantic properties. Therefore, since our project is a metaphysical reduction, the debate on the notion of function should provide us a description of a certain property (*having a function*). Our concern is not with the way scientists think, but rather with the nature of certain properties.

Secondly, it is dubious that the notion of function can play the role teleosemanticists want it to play if it strongly departs from scientific usage. For instance, we want the notion of function to ground normativity, so our notion of function should make these claims plausible: 'What a heart *is supposed to* do is determined by its function'; 'if the heart does not perform its function it *mal*functions'. But suppose we

1 Millikan (1984, 1989) defined a category that she labeled 'proper function' (Millikan, 1989,1993). What project was she pursuing when she introduced this notion?
On the one hand, Millikan has insistingly claimed that her notion of 'proper function' is not supposed to be considered a piece of conceptual analysis, but a tool for theory construction (Millikan, 1984, ch. 1; 1989, p. 290). Millikan's aim is to introduce a new concept which is supposed to help her in the naturalization of content. She needs a theoretical notion that can do a certain job in her naturalistic project and, since she thinks none of the current scientific or philosophical notions can actually play this role, she puts forward a new technical concept: 'proper function'. So, one might think, her project is of the third kind, i. e. a stipulative definition.
However, she does not intend the notion of 'proper function' to be merely stipulative (although see Millikan, 2002, p. 114). She admits that there are certain strong connections between 'proper function' and the scientific notion of 'function'. In particular, she claims that her notion of proper function is intended as a 'theoretical definition':

> However, although it makes no material difference for the uses to which I have put the definition whether it is or is not merely stipulative, I believe that it is not merely stipulative (...). A theoretical definition is the sort the scientist gives you in saying that water is HOH, that gold is the element with atomic number 79 or that consumption was, in reality, several varieties of respiratory disease, the chief being tuberculosis, which is an infection caused by the bacterium *bacillus tuberculosis.* (...) let me say that my definition of "proper function" may be read, roughly, as a theoretical definition of function. (Millikan, 1989, p.290)

So, I think Millikan's own notion of 'proper function' is probably intended as a real definition. In the next section, I will put foward some arguments why teleosemantics should indeed provide a real definition of function, rather than a nominal or a stipulative one.

introduce a technical notion N, which is not supposed to correspond to any scientific term. Why should we think N grounds normativity? Why should we think that satisfying the criteria for qualifying as an N suffices for being able to malfunction? Of course, we could also set up a second technical notion of 'malfunction', but the same problem would simply reappear.

Let me put the same point in a different way. If the notion of function teleosemantics puts forward seems to most people to be grounding normativity (and content), it is because it seems to pick up a real property we are somehow familiar with. We know that the functions we usually attribute ground normativity (see below), so the fact that the teleosemanticist notion is very similar should explain why we accept that their functions ground normativity. At the very end, the plausibility of teleosemantics depends on assuming a close connection between the notion of function and some real property identified in science and common sense.

These arguments suggest that teleosemantics should look at a real definition of function. Only a real definition (a definition of what the property *having a function* consists in) will contribute to the naturalization of semantic phenomena.

Nevertheless, let me stress that the way scientists talk and think is also important, because the property we are trying to capture is the one that scientists pick up when they use their notion of 'function'. That is, even if we are not primarily interested in what scientists think when they attribute functions, science is the best guide in order to know which are the functions of certain entities. As I said, in a large number of cases, scientists attribute functions by appealing to the features that also explain why an entity has that function. That fact should not be surprising; we should expect to find a large overlap between the properties that scientists use in order to ascribe functions and the properties that indeed make an entity to possess a certain function. The crucial issue is that our goal is not to describe a concept but to describe a property.

These reasons suggest that a satisfactory teleological theory needs to look for a real definition of function. Fortunately, I think that most philosophers have indeed been concerned with this project. So in the following discussion on alternative theories of function, I will be assuming that they are all seeking to provide a real definition of function, rather than a nominal or a stipulative definition.

2.1.2 *Function controversy*

Since the 1970s there has been an intense controversy on the notion of 'function' and, even though it is undeniable that much progress has been made on that matter, we still lack a unified theory (Wouters, 2005; Godfrey-Smith, 1993). There have been two main theories on the market, Systemic Theories and Etiological Theories, which I will discuss in some detail below. Moreover, I would like to bring forward a third recent contender, namely Organizational Theories. They have been proposed as an unifying account of function and, for reasons I will put forward below, I think they deserve a careful consideration.

Before presenting all accounts in more detail, let me set up two desiderata that any satisfactory theory of function should comply with. These desiderata derive from the previous discussion on the goals of a

theory of function and also from the way this concept is employed in science. Furthermore, they are explicitly endorsed by many different people and they have independent intuitive support. A final motivation for these desiderata (which only concern those that are interested in the teleosemantic project) is that any theory that fails to meet them will probably be unable to solve the four problems of naturalistic theories I pointed out in the previous chapter. As a consequence, any account that fails to fulfill these desiderata is a non-starter account of functions (Artiga, 2011).

It is a commonplace that functions are a particular kind of effects. The first desideratum concerns an appropriate distinction between a trait's functions and other non-functional effects of a trait:

(Acci) A trait's function is appropriately distinguished from a trait's accidental effects. (Wright, 1973; Wouters, 2005, p. 134.)

Hearts have many effects. In particular, they make thump-thump noises and pump blood. However, the heart's function is to pump blood and not making thump-thump noises. We want a suitable notion of function to be able to distinguish accidental from functional effects[2]. This desideratum directly derives from the discussion in the last section: the functions attributed by our theory should match (to a great extent) the functions attributed by science. If our notion of function is too liberal, we might be failing to describe the property scientists pick up when ascribing functions.

A second important desideratum that also comes from the previous discussion is the following:

(Norm) A trait's function determines a criterion against which the activity of the trait is normatively evaluated. (Wouters, 2005, pp. 133–134; Krohs & Kroes, 2009; Mossio et al., 2009a, p. 814)

Since hearts have a function, they can also malfunction or disfunction. Traits malfunction when they fail to accomplish their function. Thus, an account that predicts that traits can never malfunction is surely wrong. In the same way that any satisfactory theory of content has to account for the fact that representations can sometimes be false or inaccurate (see 1.2.2.2), the right theory of functions has to allow for malfunctions. It has to explain this normative dimension of functional talk.

Let us consider now the three main approaches to the notion of function and assess whether they satisfy these desiderata.

2.1.2.1 *Etiological accounts*

The main idea of etiological accounts is that the functions of devices should be identified with *reasons* for the existence of those devices (Godfrey-Smith, 1996, p. 180; Martínez, 2010). The most important contribution to the formulation of an etiological account of function was Wright's (1973), even if other people suggested similar accounts in the 1970s (see Ayala, 1970).

Wright's definition is the following:

2 Here I follow Wright (1973) in labeling non-functional effects 'accidental', even though this expression is not ideal. 'Accidental effect' seems to entail that it is an effect that could have easily been different and this is definitely not what is intended here. Accidental effects are effects that are not functions. Any effect (sprandels, causal effects, necessary effects or any other) can qualify as accidental in that sense.

EARLY ETIOLOGICAL

The function of D is F iff:

(1) D is there because it does F.

(2) F is a consequence of D being there.[3]

To a first approximation, according to EARLY ETIOLOGICAL the function of hearts is to pump blood because hearts pump blood (condition 2) and because the fact that hearts pump blood explains why hearts nowadays exist (condition 1).

There are, at least, two important virtues of Wright's account. First, it draws a clear distinction between accidental and essential effects, so it satisfies (Acci). For instance, noses have many positive effects for humans, like enabling respiration and supporting glasses, but we want only the former to be a function of the nose. EARLY ETIOLOGICAL gets this result by identifying functions with the effect that explains why the trait exists. Supporting glasses is not the nose's function because noses are not there due to this performance, i.e. 'supporting glasses' fails to fulfill condition 1 in EARLY ETIOLOGICAL. The second advantage of the account is that the function of a trait D is due to D's actions. D's own activity justifies its function and it is not parasitic on any external goal of the organism that contains it, which as we will see is a source of problems for alternative theories. So EARLY ETIOLOGICAL easily deals with two of the problems that more strongly affect other proposals.

Nowadays, most theories of function rely upon Wright's insightful account. However, two remarkable drawbacks suggest that EARLY ETIOLOGICAL needs to be amended in significant respects.

COUNTEREXAMPLES The first important objection was raised by Boorse (1976, p. 75):

> A horned buzzing on a woodshed so frightens a farmer that he repeatedly shrinks from going in and killing it. Nothing in Wright's essay blocks the conclusion that the function of the buzzing, or even of the hornet, is to frighten the farmer. The farmer's fright is a result of the hornet's presence, and the hornet's presence continues because it has this result.

More generally, the objection is that EARLY ETIOLOGICAL is too cheap, because it tends to warrant functions to devices that intuitively lack them. Arguably, the problem is rooted in the vagueness of 'because' in condition 1. To claim that 'D is there because it does F' is not restrictive enough. The performance of F can explain in many different ways the presence of D, and some of these ways may yield counterexamples to EARLY ETIOLOGICAL. If we want our concept of function to do the job it is supposed to do, we need a more specific notion. As we will see, the most common way of dealing with this difficulty is by using the more stringent concept of *being selected for*.

3 Notice that condition 2 seems to be redundant. If D is there because it does F, then obviously D does F, and so F is a consequence of D. Nonetheless, even if 1 somehow presupposes 2, each condition seems to be emphasizing a different requirement. 2 asserts that D *does* F and 1 that D doing F must *explain* why D is there. These are the key intuitions that have motivated more recent etiological accounts. Thanks to David Pineda for pointing out this possible problem.

MALFUNCTION The second worry is a bit harder to formulate because EARLY ETIOLOGICAL fails to distinguish types from tokens. What I want to show is that either if it is interpreted as referring to types or as referring to tokens, EARLY ETIOLOGICAL cannot be right as it stands. The most serious problem is that EARLY ETIOLOGICAL precludes traits from malfunctioning, so it fails to satisfy the second desideratum set up above (Artiga, 2011).

'D' refers to a token: Suppose 'D' is interpreted as referring to a token. Then, two clear difficulties seem to follow. First, condition 1 seems to be false of most traits. It is not true that my left kidney exists because *it* does anything. Proof of that is that it would still exist even if it did not perform any activity. But, secondly, if 'D' refers to a token, then EARLY ETIOLOGICAL cannot account for malfunction (so it fails to satisfy (Norm)). In a nutshell, here is the reason: condition 2 seems to be saying that D has a function to F only if D does F ('F is a consequence of D being there'); hence, if D does not perform F, then D has no function. But, typically, D malfunctions when D's function is to F and D does not perform F. So malfunction is impossible.

One might suggest that condition 2 should be interpreted as saying that F is a *usual* consequence of D being there. Prima facie, that seems to leave room for malfunctioning, since a trait can sometimes fail to do what it usually does. But there are two serious drawbacks of this proposal. First, on that proposal, you could turn an effect that constitutes malfunctioning in a functional effect by merely increasing the frequency at which D does F. But that seems wrong. Secondly, in general whether an effect is a function does not seem to depend on the frequency a trait performs it. Hearts make thump-thump noises very often, but this is not one of its functions. Similarly, the function of a bee sting is to hurt or kill a possible predator. However, a certain bee might never use the sting. Examples could be easily multiplied.[4]

'D' refers to a type: As a reply, one might argue EARLY ETIOLOGICAL avoids this worry if 'D' is interpreted as type. But if 'D' refers to a type, condition 2 requires further clarification. If we interpreted 'D' as a type, what could justify the claim that, for a given D and F, 'F is a consequence of D being there'? I can think of two possible proposals: it might be that 'F is a consequence of D being there' is true when most Ds perform F or when Ds are able to do F. None of this proposals solves the problem of malfunction.

First, one might argue 'F is a consequence of D being there' is true when most tokens of D do F. However, this proposal has the same problems as the previous appeal to frequency. On the one hand, if that were the right interpretation, you could turn something that is not functional into something functional by merely increasing the number of entities that have this effect. But that seems wrong. As Neander (1991) puts it, you cannot health a disease by spreading it. Even if most kidneys malfunctioned, its function would remain the same. More generally, the worry is that function attributions do not seem to depend on statistical frequency. Millikan (1984) provides the example of spermatozoa; arguably, the spermatozoa's function is to fertilize an

4 Remember that condition 1 cannot help to solve this worry, because if D is interpreted as a token, 1 is false.

ovum, but in fact only one in millions actually achieve this goal. Having a function does not seem to depend on what most members do.

Wright considered a different solution in the following passage:

> In some cases we will allow that [D] does F even though F *never* occurs. All that seems to be required is that [D] be *able* to do F under the appropriate conditions (Wright, 1973, p.58. Emphasis added).

However, this proposal does not seem to solve the worry. Even malformed hearts that are unable to pump blood have this activity as their function. So function attributions cannot merely rely on tokens of D being *able* to do F.

In conclusion, if D is interpreted as a type, it is not clear what makes true the claim that 'F is a consequence of D being there'. Neither the fact that most Ds do F nor the fact that Ds are able to do F seem to do the trick.

In conclusion, neither if 'D' is interpreted as a type, nor if it interpreted as a token can we account for the desiderata. Fortunately, the good news for the etiologist is that he can solve the problem by appealing to a type/token distinction and the property *being selected for* (we will see that systemic accounts cannot use this solution). Even though a particular d (token) does not perform F, d can have the function of F iff (1) d belongs to the type D and (2) Ds have been selected for F. Condition 1 crucially appeals to a type/token distinction and condition 2 will be defined by adding a historical condition: natural selection.

Let us sum up the results of this section. Wright's account is very promising, but I have identified two important drawbacks: the notion of 'because' is too imprecise and it cannot account for malfunction. I suggested that both problems can be solved by appealing to the notion of *being selected for* (and making a type/token distinction). Introducing these elements, however, will require a precise definition of some some notions and a significant modification of Wright's account. This is the task of the next section.[5]

2.1.2.2 *Types/tokens, Darwinian Populations, Selection for and Etiological Functions*

In the last section, I claimed that there is a surprising connection between the two main problems of Wright's account: both can be solved by appealing to the notion of *being selected for.* In this section I will argue there is another interesting link between two other concepts: despite appearances, the question of establishing a type/token distinction and the definition of *being selected for* are very closely intermingled (at least, in the context of our present discussion). The goal of this subsection is to define more precisely these two notions (and some others) and specify their intimate relation.[6]

REPRODUCTIVELY ESTABLISHED FAMILY First of all, we need to explain what it is for a particular trait d (token) to belong to a type D.

5 Let me point out that, while many philosophers working on the notion of function appeal to the process of selection and also find Wright's account unsatisfactory, the precise connection between these two ideas has not been developed in the literature.

6 Again, I would like to stress that thess connections between the problems of Wright's account, then notion of *selection for* and the type/token distinction have been largely overlooked in the literature.

Here it will be useful to appeal to Millikan's notion of a reproductively established family (Millikan, 1984, ch. 1):

REPRODUCTIVELY ESTABLISHED FAMILY A group of individuals $d_1, d_2, d_3, \ldots d_n$ form a reproductively established family D iff $d_1, d_2, d_3, \ldots d_n$ tend to resemble each other in important ways because they are the result of some causal process of copy.

In short, the idea is that a reproductively established family (REF) is a group of individuals that tend to have some properties in common because they have been copied (Reproductively Established Families will be further specified in 3.2.4).

The notion of reproductively established family (REF) applies to many entities: The car *Ford Fiesta*, the species *Canis Lupus* or the action of shaking hands are examples of REF (Millikan, 2005). All members of these families tend to resemble the others in virtue of some causal process of copy. Here I would like to focus on the particular case that mostly concerns us: traits. Traits (the heart, the lungs, the brain,..) form reproductively established families. Lungs, for instance, tend to have many properties in common because they are the result of a process of copy. The complicated process of biological reproduction partially explains why my lungs resemble the lungs of my ancestors.

Having defined the notion of REF, we are now in a position to explain what it is for a particular trait d to belong to a type D. A trait token d belongs to a type D when d belongs to a reproductively established family D. That is, d is a D because d resembles a group of members that form D in virtue of the fact that there is an underlying process of copy. Particular traits (John's heart) belong to certain kinds (the kind *heart*), because there is an underlying process of copy that explains why this particular trait resembles other traits of the same type (why John's heart resembles the heart of his ancestors) (Millikan, 1984). The notion of Reproductively Established Family enables us to group entities into kinds and specify in more detail why a certain token belongs to a certain kind.

Now, if we focus on biological entities such as traits, it seems we can specify in more detail the kind of process of copy that takes place, which accounts for the fact that REPRODUCTIVELY ESTABLISHED FAMILY applies. Traits (and many other biological entities such as species) form what Godfrey-Smith (2009) calls 'Darwinian Populations' (I will follow him in the use of this term). To a first approximation, Darwinian Populations are reproductively established families that undergo a process of selection (such as natural selection). More precisely, Darwinian Populations have three crucial features that make them ideal for our present concerns: (1) Since they are reproductively established families, they ground a distinction between types and tokens that was missing in Wright's account; (2) Darwinian Populations are entities which undergo processes of natural selection, so the notion of Darwinian Population will enable us to make more precise the notion of 'because' of premise 1 in EARLY ETIOLOGICAL and to account for malfunction; (3) Most of the entities we will talk about in this dissertation form Darwinian Populations. So let me describe in more detail what is a Darwinian Population and how it yields an etiological definition of function that can overcome the problems of Wright's EARLY ETIOLOGICAL.

DARWINIAN POPULATION As I said, Darwinian Populations are a special kind of reproductively established family. The main feature

that distinguishes them from the rest of REFs is that they undergo a selection process. So, in order to define Darwinian Populations we need to get into a description of the process of natural selection.

Natural selection is a process that involves certain causal patterns in or between populations. In order to assess whether a certain trait has undergone a process of selection, we consider the distribution of this trait in a given population at different times and the causal process that took place between them. The idea that natural selection takes place within causally connected populations suggests that processes of selection concern reproductively established families (REFs), rather than individuals. REFs, rather than individuals, are selected. As de Queiroz Says, 'lineages are the things that evolve' (1999, p.65).

Thus, as a first approximation, by 'Darwinian Population' I will mean a certain REF that is subject to a process of natural selection. More precisely, if we rely on standard characterizations of natural selection, a Darwinian Population can be defined as follows:

Darwinian Population

D forms a Darwinian Population only if the following conditions are met:

(a) *Replication*: Members of D must form a reproductively establish family, in accordance with Reproductively established family

(b) *Variation*: The replication of members of D included some changes in some of its members.

(c) *Environmental interaction*: The interaction of members of D with certain environmental circumstances determined differential replication among its members.

Several things must be said about this definition. First of all, Darwinian Population only specifies a set of necessary (and non-sufficient) conditions for a population to be Darwinian (Godfrey-Smith,2009). The reason is that there typically are further requirements for natural selection to take place.[7] For instance, variations have to take place within a certain ratio; otherwise, they would produce nonviable individuals. Furthermore, the environmental pressure has to remain within certain limits, since a too strong selection pressure might lead a population to extinction (Sterelny and Griffiths, 1999). Cases of selection might also require that the variation responsible for the selectional outcome be of internal properties of the item that forms a Darwinian Population. This condition is intended to avoid situations as the following: what primarily explains that a given trait increases its presence in a population is that members of the same population with a different trait have been struck by a lightning bolt. Godfrey-smith adds that a further requirement for some populations is that the fitness-landscape

7 As Godfrey-Smith (2009, p. 39) suggests: "A Darwinian population in the minimal sense is a collection of causally connected individual things, in which there is variation in characters, which leads to differences in reproductive output (differences in how much or how quickly an individual reproduces) and which is inherited to some extent. Inheritance is understood as similarity between parents and offspring, due to the causal role of the parents. (...)
Any Darwinian population will have these properties plus others. The behavior of a system is determined by the particular forms that variation, heredity and fitness-differences take along with other features of the population."

be rather smooth. He is appealing here to Wright's concept of a "fitness landscape", which pictures the relationship between variation and fitness. A "smooth" landscape is one where similar phenotypes are associated with similar fitnesses; a "rugged" landscape is one where similar phenotypes are associated with very different fitness values. He argues that "smoother" landscapes are more "Darwinian" in character, presumably because it is easier to traverse them via selection. In other words, if *any* internal change (e.g. a gene mutation) can easily get into fixation, then we might say that the fact that this particular variation exists nowadays is not due to natural selection. An extreme case of rugged landscape constitutes evolution by genetic drift (Godfrey-Smith, 2009, ch.3).[8]

Thus, a complete definition of Darwinian Populations can be very complex and entirely depends on scientific investigation. Biologists are in charge of establishing the precise criteria for a reproductively established family to form a Darwinian population. At the very end, what Darwinian populations are is an empirical question. I have set up the most important conditions a REF must satisfy in order to qualify as a Darwinian Population, which are usually regarded as the fundamental or paradigmatic conditions for a selection process to take place (Sterelny and Griffiths, 1999), but it is worth stressing that the list could be extended very much (Godfrey-Smith, 2009). That should not prevent our project from getting started, since in this dissertation we are going to focus on clear cases that plausibly satisfy all known conditions for qualifying as Darwinian Populations.

That leads to a second important point. In contrast to simplified versions of natural selection (e.g. Hull et al, 2001), whether a population counts as Darwinian or not is a matter of degree. Thus, we can distinguish paradigmatic cases from marginal cases of Darwinian Populations (Godfrey-Smith, 2009, p. 41). What determines whether a given Darwinian population is a paradigmatic case depends on the degree in which it satisfies the conditions set up above, taking into account that some conditions might be more important than others. Consequently, some Reproductively Established Families are better examples of Darwinian Populations than others. Paradigmatic examples of Darwinian Populations are zebras or hearts.

SELECTION FOR In the previous section I defined a Darwinian population, that is, a REF that has undergone a processes of natural selection. This definition kills two birds with one stone: on the one hand, it enables us to specify in more detail the type/token distinction in the context of biological entities and, on the other, it provides us with the key elements for the definition of *being selected for* (Sober, 1984):

SELECTION FOR

D is *selected for* F iff:

1. D forms a Darwinian Population, in accordance with DARWINIAN POPULATION.
2. F is an effect of some members of D.

8 Godfrey-Smith (2009) thinks that there is a logical continuum between processes of natural selection and evolution by drift. Drift constitutes, so to speak, very marginal cases of natural selection (so marginal, that in fact they form a category *sui generis*).

3. F is the effect that (in a preponderant number of cases) causally explains why differential replication favored members of D that could do F.[9]

As previously noted in the context of Darwinian Populations, it must be kept in mind that there are different degrees in which a certain causal process taking place within an organism might satisfy Selection-for (Godfrey-Smith, 2009). For instance, sometimes there is no definite answer to the question whether certain members of D had the function to F, because it might not be clear whether there has been enough previous cases that causally explain differential replication. Hence, *being selected for* is a vague property.

As a result, there are at least two independent sources of vagueness. On the one hand, whether a given REF forms a Darwinian Population is a matter of degree; on the other, whether a given member of a REF has been selected for is also graded.

etiological function Let us go back to the main topic of this section. We saw that we need to improve Wright's definition early etiological in two crucial respects. We required a principled way of distinguishing types from tokens and a precise formulation of being selected for. I have spent some time defining the relevant notions of reproductively established family, Darwinian Population and selection for, so I think we are now in position to present a better etiological account of function:

Etiological Function

A trait d has the function F iff:

1. d is a member of D.
2. D forms a Darwinian Population, in accordance with Darwinian Population.
3. D has (recently) been selected for performing F, in accordance with Selection For.[10]

Notice that in 3 I included a further condition: in order for a trait D to have a function, D must have *recently* been selected for.This temporal condition is required in order to avoid attributing functions to vestiges, which are traits that served a function in the distal past but have not recently been selected for or selected against (Griffiths, 1993). The standard way of dealing with these cases is by introducing a temporal condition, such that a trait has a function to F only if it has recently been selected for F. That turns the etiological account of function into what some people call a 'modern history approach' (Godfrey-Smith,1994; Martinez, 2010).

The paradigmatic case that exemplifies Etiological Function is the heart: the function of my heart is pumping blood because (1) my heart is a member of the kind *heart* (i.e. is a member of the REF *heart*) (2) the kind *heart* is a Darwinian Population (3) hearts have recently been

9 As Sober (2010) wrote: "If there is natural selection for pumping blood, this means that pumping blood causes enhanced survival and reproductive success."

10 It could be sensibly argued that condition 2 is entailed by 3, because according to Selection for only Darwinian Populations can be selected for. This is entirely right, but I decided to add condition 2 in order to make explicit that any functional trait d must be a member of a Darwinian Population. This is central aspect of the etiological theory that will have important consequences in this dissertation (for instance, in 3.2.6).

selected for by natural selection because they pumped blood (and not, for instance, because they made thump-thump noises).

It is not hard to see that Etiological Function can easily overcome the problems of Early Etiological. First, it can avoid clear counterexamples, like the hornet's buzzing, because this is not a case of natural selection that satisfies the conditions specified in Darwinian population. Secondly, Etiological Function adequately distinguishes types from tokens: my heart has a function of pumping blood in virtue of belonging to a Darwinian Population (a special sort of reproductively established family that has been selected for pumping blood). Thus, my heart malfunctions when it does not produce the effect that hearts have been selected for producing. Therefore, Etiological Function can also satisfactorily account for malfunction.

Let me stress that Etiological Function keeps the key intuition of etiological approaches (such as Wright's or Ayala's), namely that functions are reasons for existence. The heart's function is to pump blood because hearts were selected for pumping blood. And since hearts were selected for this performance, pumping blood partially explains why hearts nowadays exist. Thus, if a trait has a function F, F (partially) explains why this trait exists.

I have already pointed out some advantages of etiological theories over its competitors, specially the fact that it fulfills (Acci) and (Norm). But, before moving forward, let me point out its most important difficulty. One of the consequences of Etiological Function is that a trait's function depends on the performances of past traits of the same type. In other words, what a particular trait token does or is able to do is completely irrelevant in order to attribute functions to it; only past actions performed by its ancestors matter. To some people, that sounds counterintuitive (Mossio et al, 2009a, 2009b; Wouters, 2005). More importantly, as we will see in the next chapter, this intuition is the source of the Swampman problem for Teleosemantics (see 3.3.4). We will have the chance to discuss these consequences in more detail in the next chapter.

Let us now move on to an alternative account of functions, often called 'Systemic view'.

2.1.2.3 *Systemic Accounts*

Historically, systemic views appeared roughly at the same time as etiological views. The key tenet of these theories is that a function of a device is determined by the contribution it currently makes to a system. Very roughly, their claim is that if a particular trait d contributes to the system by performing F, then d's function is to F. Crucially, notice that this is precisely the intuition that tells against etiological views. According to etiological theories, what explains that a heart (token) has a certain function is the fact that it belongs to a certain type that has been selected for a certain task. The idea that functions of particular traits depend on facts far-removed from the actual situation is taken by many to be the main problem of the etiological view (Wouters, 2005; Mossio et al. 2009a), and this weakness is precisely taken as the starting intuition for systemic views.

Independent motivation for systemic views can be found, for instance, in artifacts; if we want to know the function of the carburetor, we just need to look at the contribution it makes into the motor engine. Similarly, it has been suggested, for the biological world. Biologist

seems to discover the function of a trait by paying attention to what it does in the organism. The key insight, hence, is that functions are effects that devices have within systems. This is why they are usually called 'Systemic accounts'.

Whereas all systemic approaches hold this key intuition, there is little consensus as to what is the kind of contribution that determines function. Some proposals have been the following: something useful (Canfield, 1964), good (Sorabji, 1964) or, more generally, contributions to goals of the system (Boorse, 1976). So we can summarize the main idea of these views in the following definition:

SYSTEMIC

A function of an item d in a system S is to do F iff:

(1) d does F in S

(2) If d did not do F the life conditions (pleasure, fitness,...) of S would be worse.[11]

There are, at least, three important problems with this account.

The first reason against adopting SYSTEMIC has to do with our naturalistic project, rather than with any incoherence of the account. The worry is that there is no independent way of choosing one among the different systems of reference S (Millikan, 1993). The system that determines the function of the heart, for instance, is the circulatory system or the human body-plus-doctor-stethoscope? Crucially, note that different systems may attribute different functions to a given trait. If the system of reference is the circulatory system, then the heart's function might be to pump blood, but if the system is the body-plus-doctor-stethoscope, then making thump thump noises can also be considered a function of the heart; after all, making thump thump noises increases the life expectancy of the body-plus-doctor-stethoscope (say, by increasing the human's life expectancy). Since systems can be identified *ad libitum*, it seems functions are multiplied without restrictions. Cummins' (1975) reply to this sort of problem is that function attribution depends on the system that *we take* as relevant for our purposes. Hence, to avoid an excessive attribution of functions to any device, SYSTEMIC needs to appeal to the intentions of the observer in order to pick out the right system of reference. Since that would introduce an unanalysed intentional state within the φ-facts, a systemic account can not do the job it is supposed to do in a naturalistic project. Of course, this is not an argument against the claim that SYSTEMIC is the right theory of functions; it just points out that if we were to agree with systemic theorists, teleosemantics would be doomed.

Secondly, SYSTEMIC fails to distinguish accidental from functional effects (Wright, 1973). According to SYSTEMIC, both enabling respiration and supporting glasses are functions of noses, because both are effects that contribute to the life conditions of the system (the organism). Indeed, every positive effect that device confers to the system would be considered a function. But that is at odds with common sense and scientific usage. In that sense, SYSTEMIC is too weak because it attributes too many functions. Here the problem lies in condition 2 of SYSTEMIC.

11 As I said earlier, I formulate the systemic view as an account of biological items, because this is the kind of entities we will be concerned with in this thesis. However, a systemic account of *artifactual* functions would be spelled out in a different way (e.g. it would not appeal to 'life conditions') and would probably require a different kind of discussion.

Thirdly, Systemic relies too strongly on the actual state of the device and for this reason it cannot account for malfunction. Imagine that someone has a malformation in the kidney, such that it can not filter wastes from blood any more. In this case, the kidney does not contribute to any goal of the system, but we still want to say that its function is to filter wastes from blood. Unfortunately, if a trait does not have a function, it can not malfunction either. Thus, Systemic is too restrictive because it fails to attribute functions to traits that intuitively should have them. That shows that this account fails to satisfy (Norm). This problem is rooted in condition 1 of Systemic.

In that respect, and in contrast to etiological theories, Systemic cannot introduce a type/token distinction or appeal to natural selection in order to solve these worries; if they did, they would turn Systemic into a special sort of etiological theory. Let me explain.

type/token and malfunction Let us try to introduce a type-/token distinction and see whether in this way Systemic can account for malfunction. Suppose we add a condition 0 which states that d is a member of D, a condition 1 which states that the system s (token) is a member of systems S (type), and replace d for D in 1 and 2 accordingly:

Systemic*

A function of an item d in a system s is to do F iff:

(0*) d is a member of D

(1*) s is a member of S

(2*) D does F in S

(3*) If D did not do F the life conditions (pleasure, fitness, ...) of S would be worse

On a first approximation, Systemic* seems to be in a position to solve the problems of Systemic. First of all, it can attribute a function F to a trait d even if d does not perform F or even if d is unable to perform it. On Systemic*, particular traits have functions in virtue of belonging to a type D, which satisfies certain conditions. Secondly, it seems that Systemic* still retains the broad intuition motivating systemic accounts; functions are contributions to goals of a system. Therefore, Systemic* is a clear improvement over Systemic.

However, now 2* becomes problematic. What justifies the claim that 2* is true, that is, that 'D does F in S'? We saw that interpreting 'D does F' as *most* members of D do F faces difficult problems. On the one hand, the spermatozoa's function is to fertilize an ovum, but very few of them actually achieve this goal. If we adopted this first reading of 2*, spermatozoa (type) would never have this function, since most spermatozoa (tokens) fail to do F. Similarly, we also wanted to keep the intuition that even if most kidneys malfunctioned, its function would remain the same (remember Neander's dictum: we can not health a disease by spreading it). So what could make true condition 2*, i.e., that 'D does F in S'?

An option is to change 2* by 2**, which claims that 'D *is supposed* to do F in S'. But, again, how are we to explain what D is *supposed to* do? Of course, we could say that, according to the design of S, D *should* perform F. This idea easily makes sense in intentional design

(i.e. artifacts), but how can we explain what a device is designed to do in biology? The only possible way of accounting for design in the biological world is by appealing to natural selection (Dennett, 1996). If a trait has been selected for F, we can say that a trait is supposed to F, in accordance with design. Only natural selection can account for what a trait-type is supposed to do, so 2* should be substituted by 'D has been selected for F in S'.

But notice that introducing natural selection has a striking consequence in Systemic*. Once we appeal to the selection of a trait, condition 3 (and 3*) becomes unnecessary. A trait has been selected for F, if F is the effect that explains why traits of that type exist. Hence, the claim that a trait has been selected for F assumes that F has been fitness-enhancing for some system (individual or group), so we do not need to specify in the definition to which system the trait contributes. That amendment not only avoids duplications in the definition, but it also solves the problem of specifying the relevant system in non-intentional terms. As a result, we obtain the following definition of function:

Better Systemic

A function of an item d is to do F iff:

(0) d is a member of D

(1) D has been selected for doing F

Certainly, the appeal to natural selection allows Better Systemic to solve the problem of previous versions of the theory. Unfortunately, if we appeal to natural selection, we are abandoning the key insight of systemic views, namely that a trait's function depends on the trait's current contribution to a system. In other words, better systemic does not look like a systemic account at all. Indeed, it is not hard to see that better systemic just is an etiological account of function. The upshot is that the only way of improving better systemic, in a way that be able to account for (Acci) and (Norm), is by adopting an etiological account.

2.1.2.4 *Organizational Accounts*

During many years the debate on the notion of function was dominated by systemic an etiological theories. However, there has recently appeared a third contender that deserves careful consideration: organizational theories. The chief innovation of these accounts consists in putting forward a new element: self-maintained systems. Proponents of this innovative approach claim that the notion of self-maintained system is a well-established concept in some scientific disciplines, such as biology and thermodynamics (Mossio, et al. 2009a). That is especially important in this context, because a common goal of functional accounts in philosophy is to account for functions as they are attributed in science (see 2.1.1). Hence, a category that has been borrowed from these sciences obviously has the credentials for taking part in this project.

Before presenting the organizational definition of function, there are two key notions that should be introduced: organizational closure and organizational differentiation.

Organizational Closure A system is *Organizationally Closed* when there is a circular causal relation between some macroscopic

> pattern or structure and its microscopic dynamics and reactions (Mossio et al. 2009a, p.824).

Basically, the idea is that organizationally closed systems are characterized by a feedback mechanism, which consists in the fact that the system's effects help to regenerate the parts of the system that produce or maintain these effects. A clear example is a candle flame. Some microscopic reactions of combustion generate the flame, and the flame in turn contributes to the existence of the microscopic reactions that generate and maintain the flame alive.[12] Notice that in organizationally closed mechanisms, the activity of the system is a necessary but not sufficient condition for the maintenance of the system. For example, in the case of the candle, oxygen and wax are also required for regenerating the flame.

While exhibiting organizational closure is supposed to be a necessary condition for a system to have functional parts, it is surely not sufficient. Otherwise, whirlpools and candle flames would ground function attributions. What we need, according to people endorsing this view, is organizational differentiation:

> Organizational Differentiation: A system is *organizationally differentiated* when it is possible to distinguish parts that contribute in different ways to the self-maintenance of the system.

For instance, the idea is that hearts, livers and lungs are differentiated parts of the body that help to regenerate the organism in different ways. Since the human body has many different parts in this sense, we can say it exhibits organizational differentiation. In contrast, if every part of the system contributes in the same way to its reproduction, then the system is *not* organizationally differentiated.[13]

Systems which are closed and differentiated are what people working in this tradition call 'self-maintained systems'. The assumption that there are closed and differentiated systems is what enables these philosophers to put forward an original definition of function (McLaughlin, 2001; Christensen and Bickhard, 2002). The most specific version of the organizational definition we have (due to Mossio, et al., 2009a, 2009b) is the following:

> Organizational function
>
> A trait T has a function iff:
>
> (1) T contributes to the maintenance of the organization O of S.
>
> (2) T is produced and maintained under some constraints exerted by O
>
> (3) S is organizationally differentiated

In other words, a trait T has a function if, and only if, it is subject to organizational closure in a differentiated self-maintaining system S.

12 Mossio et al. claim candle flames and Bénard cells are examples of what they call 'dissipative structures'. In order to avoid unnecessary terminological complexities, I will avoid presenting terms that do not make any substantive contribution to the theory.

13 Mossio *et al.* (2009a, p. 826) claim that for a system to be organizationally differentiated, another condition that needs to be met is that their parts should be created or maintained by the system.

A first advantage of Organizational function over Systemic is that in the former we have an independent way of picking out the relevant system of reference. While Cummins accepts that function attributions are observer-relative (i.e. they depend on the system we take as reference), Organizational Theorists provide a principled way of picking up the relevant system by appealing to self-maintained systems, which satisfy Organizational closure and Organizational differentiation. Furthermore, as we pointed out earlier, it relies on certain notions that have been extensively defined in science, so it has the right credentials for figuring in a naturalistic project.

Despite these advantages, Organizational function is not devoid of problems. Barring certain technicalities[14], there are (at least) three serious difficulties with this approach.

The first objection to Organizational Theories is that they fail to draw a distinction between accidental and functional effects, and hence fail to accommodate (Acci). For instance, since my nose supports my glasses and helps my self-organized system (i. e. me) to survive better, one should conclude from Organizational Function that one of the functions of my nose is to support glasses, which is clearly counterintuitive.[15]

The second problem shows up when trying to state what it is for a trait to malfunction. According to the supporters of Organizational function:

> Dysfunctions appear whenever a trait fails to adequately perform its primary and/or secondary function. A dysfunctional trait is a trait that fits 2 and 3 but fails to fit 1. (Mossio et al., 2009a, p.833)

The main reason why this idea cannot work is that Organizational function establishes the conditions for function *possession*, not for a trait's *fulfilling* its function. Organizational function states the criteria that any trait must comply with in order to *have* a function. From that, it follows that if a trait does not fit 1, it would not have a function. And, of course, if a trait does not have a function, it cannot malfunction either. Therefore, if we accept Organizational function and a trait does not satisfy one of the conditions of Organizational function, then a trait lacks a function, which is to say that it cannot dysfunction.

As we saw, this objection is familiar to Systemic Accounts and affects any theory that bases function attribution on some contribution to the

14 One specially notorious problem with Organizational function is that it specifies the conditions required for a trait to have *a* function, but it cannot establish *which* is this function. In other words, if a trait satisfies Organizational function, we can conclude it has a function, but we have no clue as to what it is. Nonetheless, this problem can easily be solved in the following way:
A trait T has the function *to F* iff (1) T contributes *by F-ing* to the maintenance of the organization of S (2) T is produced and maintained under some constraints exerted by O (3) S is organizationally differentiated. For an extended discussion of these and other problems, see Artiga (2011).

15 In fact, this counterintuitive conclusion becomes a real problem once we take into account the fact that function attributions are also supposed to explain why the trait exists, that is:

> (...) all functional attributions to a trait T, be they primary or secondary, provide an answer to both the question 'why T?' and the question 'what is T for?'. (Mossio et al., 2009a, p. 832)

So Organizational function entails that the fact that noses support glasses explains why noses exist. This result is clearly unsatisfactory.

system. On these accounts the distinction between having a function and failing to fulfill it collapses. Again, one could try to explain malfunction by introducing a type/token distinction; if one takes this option, having a function would depend on belonging to a type and performing a function on actually carrying out the activity that a trait is supposed to perform. But then, of course, the question is what determines that a given type is supposed to do F. We saw that this cannot be determined by the fact that most members of D perform F; the only reasonable answer seems to be that members of type D are supposed to perform F because members of D have been selected for F. However, if ORGANIZATIONAL FUNCTION takes this option, it will end up as a sort of etiological theory. In contrast. remember that ETIOLOGICAL FUNCTION does not have this problem. According to ETIOLOGICAL FUNCTION, a trait token malfunctions when it does not perform the effect that explains why traits of its type had been selected for.

The final problem of ORGANIZATIONAL FUNCTION is that it has counter-intuitive results concerning cross-generational traits.[16] Cross-generational traits are traits that do not contribute in any relevant way to the organisms carrying them, but only to organisms that belong to the next generation. For instance, sperm does not contribute in any important way to the organism that carries it. Since, according to ORGANIZATIONAL FUNCTION, a trait has a function only if it contributes to the system that contains it, cross-generational traits do not satisfy 1 and hence lack a function. Notice that the same problem can be extended to many other traits whose main effects concern organisms in the next generation. For instance, teaching or caring behaviors from parents to offspring would lack a function according to ORGANIZATIONAL FUNCTION.

A not very satisfying way of dealing with this problem is by adopting a 'splitting account', according to which there are two kinds of functions (Delancey, 2006). The claim that cross-generational traits have a special kind of functions, which is different from the rest of functional traits, is clearly unappealing. Are really cross-generational traits *so* special? It seems sperm and caring behaviors have functions in exactly the same sense lungs and self-defensive behaviors do.

A more interesting reply has recently been suggested by Mossio et al. (2011), who appeal to 'encompassing systems'. An encompassing system is a system composed by a set of members of lineage. Now, according to Mossio et al. (2011, p. 200) these encompassing systems exhibit organizational closure and organizational differentiation:

> (...) the organization of the 'encompassing system' composed by a reproducer and a produced system itself fits the characterization of a self-maintained system. The process of reproduction, in this sense, simply constitutes one of the functions through which the organization succeeds in maintaining itself beyond the lifespan of individual organisms. Since the encompassing system composed by producer and reproduced organism possesses a (temporally wider) self-maintaining organization, reproductive traits are subject to organizational closure, and their functions are correctly grounded in the organizational account.

16 The section on cross-generational traits has been developped with the help of Manolo Martinez.

The idea is that, after all, there is an organization that underpins the attribution of functions to cross-generational traits, namely the encompassing system, which consists of a set of different members of the same lineage. Sperm, for instance, contributes to the self-organization of the encompassing system, which includes a set of organisms of different generations (and also the organism carrying the sperm). At the same time, the encompassing system (the lineage) creates and maintains the sperm in later generations, and hence conditions 1, 2 and 3 of Organizational function seem to be fulfilled. So, if encompassing systems qualify as self-maintained systems, Organizational function can attribute functions to cross-generational traits in the same way it attributes them to the rest of functional items.

The problem with this reply is that if organizational theorists adopt this view, the Organizational Account turns into an etiological theory (again). Consider, for instance, what grounds the attribution of function to rabbit sperm. Rabbit sperm has a function because:

1 Rabbit sperm has contributed to the maintenance of the encompassing organization of the rabbit lineage.

2 Rabbit sperm is maintained and produced by the encompassing organization.

And, of course, for Rabbit sperm to have a function, this causal loop has to take place within several generations (Mossio et al., 2011). The upshot is that the sperm of a particular rabbit has a function in virtue of the fact that (1) it has been produced by the lineage (that is, it has been copied from previous members) (2) the sperm of his ancestors have contributed to the reproduction of (a significant number of members of) the lineage. I think it is pretty obvious that this definition is very close to Etiological function.[17]

In conclusion, it seems that the only way Organizational function can account for normativity and cross-generational traits is by adopting some kind of etiological account of function. On the one hand, it can explain malfunction only if a type/token distinction is made, but the only way of providing such a distinction is by turning the view into an etiological theory. On the other, the problem of cross-generational traits also forces Organizational function to appeal to encompassing systems (lineages) and hence functions depend on the contribution a trait makes to the existence of the lineage. That solution also turns Organizational function into a version of an etiological theory. In

17 The problem of the functions of cross-generational traits has brought the connection between Organizational function and etiological accounts of function into focus, but we can now see that the link between Organizational function and etiological theories was already present before any appeal to cross-generational traits was made. The reason is the following: Organizational function attributes a function to a trait when there is a causal loop between the trait and the organization whence it belongs. The former contributes to the self-maintainance of the latter and, in turn, the latter produces and maintains the former. Now, crucially, for that causal loop to take place, a certain period of time is required. Saborido et al. admit that this is a conclusion of their account:

> The first remark is that a self-maintaining organization occurs in time, and can be observed only in time. Thus, ascribing functions to traits or parts requires the consideration of a system that realizes self-maintenance during a period of time long enough for organizational closure to be observed. (Saborido et al., 2011)

So, even before appealing to encompassing systems, Organizational function was attributing functions in virtue of *past* performances of a trait. All along, there was an appeal to historical reasons for the existence of the trait built into Organizational function.

a parallel fashion, we saw that Systemic views can only account for malfunctions if they adopt an etiological view. The result, I think, is pretty clear: we should accept an etiological view on functions.

2.1.3 *Conclusion of the Discussion on Functions*

We have seen that the Etiological Function is clearly preferable over Systemic and the Organizational Function. For one thing, the former seems to be able to accommodate the most important intuitions elicited by function attributions: it distinguishes functional effects from accidental effects and it accounts for the normativity involved in function attributions. The fact that only etiological theories satisfy these desiderata strongly suggest that they rightly capture the property of *having a function* that is commonly attributed in science. Additional support will be provided in part II of this thesis, in which I will spell out in more detail the connections between Etiological Function and actual scientific practice.

Thus, I will assume in what follows that functions are etiological functions, as defined in etiological function. Let us see now how this concept can ground a naturalistic account of representation and content.

2.2 representational systems

Now that we have a promising definition of function, the next question we need to tackle is how it can be used in order to yield a plausible theory of representation and content. We will see that, even if the notion of function is well defined, there are many questions that still need to be resolved: it is not obvious which entities are endowed with functions, which are their functions and how they determine content. Indeed, I think that a common mistake within the teleosemantic literature is to assume that once the notion of function is in place, a satisfactory naturalistic account of representation is straightforward. On the contrary, there are many different teleosemantic accounts that can be put forward with a single notion of function. Standard teleosemantic theories have failed to address this question directly.

2.2.1 *Crude Teleological Account*

A first intuitive way of developing a teleological theory of content is to substitute the notion of causation or indication in Crude causal account or Relative indication (see 1.2.2.1 and 1.2.3.3) by the notion of etiological function. For instance, we can imagine a Crude teleological account of the following sort:

Crude Teleological Account

R represents S iff

1. R has a function, in accordance with Etiological Function
2. The function of R is to S

It is not hard to see why this account can not work as it stands. A first difficulty is that, prima facie, it seems to attribute a function to the wrong kind of entity. As it is usually understood, *traits* (hearts,

livers) or *devices* (thermometers, barometers) have functions and *states* have content. So the claim that the thing that does the representing is the same as the item that has the function, as suggested in Crude Teleological Account seems to be a conceptual confusion. If concepts have functions, then it seems that something else (a different *state*) is what has the content.[18] I do not think this is a knockdown objection since, after all, one might argue we sometimes attribute functions to states as well (see 3.2.6). My point is rather that, if we are willing to attribute functions, the natural place to look are devices rather than states.[19]

A more serious drawback is that Crude Teleological Account looks implausible as an account of descriptive content (see Stegmann, 2005, p. 1019). One could grant, for instance, that perceptual states have many functions, like causing the belief system to generate certain beliefs, enabling certain actions,... but they seem to represent something entirely different, namely the presence of certain worldly affairs. Indeed, I will argue later (see chapter 4) that the content of perceptual states is something like *there is an object with such and such properties*; however this is clearly not one of its effects. The same problem can be put in more general terms: functions are effects, but, typically, a state's descriptive content is not any of its effects. Consequently, Crude Teleological Account suffers from the Adequacy problem and, hence, it is probably wrong with respect to descriptive content.

Nonetheless, notice that Crude Teleological Account has some plausibility as an account of *imperative* content. An illustrative example is the case of beavers; beavers splash the water with their tail so as to warn their fellows that a predator is approaching. If Crude Teleological Account is interpreted as a theory of imperative content, then it entails that the (imperative) content of the beaver's splash is something like *'go hiding!'*, since this is an effect of those states. Some people think this is a plausible result and that, generally, imperative content should be identified with certain effects of representational states (Millikan, 1995, 2004; Papineau, 2003).[20] In this thesis I will focus on descriptive content, so I will leave this discussion aside.

2.2.2 *Early Papineau*

There is a second (more promising) way of adding the notion of function to a naturalistic theory of content. We just saw that the intuitive problem with Crude Teleological Account as a theory of *descriptive* content is that the representational content does not seem to be an effect of a state. Indeed, a reasonable idea is that the fact that a state has a given content explains why it has certain effects on other devices. If we push this suggestion a bit further, we get to the hypothesis that there is a

18 Notice that a type-token distinction is not what is required here. For instance, one might claim that my concept-token has content C because it belongs to the concept-type with function F. However, if concepts-type have functions and my concept-token belongs to this type, then my concept-token has a function (not content). Appealing to a type-token distinction can help you to explain why concepts-token have functions given that concepts-type have functions, but not why the same thing that has functions does the representing.

19 Here I strongly disagree with Elder (1998, p. 351), who claims that the difference between attributing functions to states or to mechanisms 'is more a matter of expression than doctrine'. For an extended discussion, see 3.2.6.

20 In fact, something like Crude Teleological Account is probably contained in the Pushmi-Pullyu account of simple representations (Millikan, 1995, 2004)

fundamental relation (other than identity) between a state's content and the effects it has on other devices. So, an alternative proposal about origin of descriptive content introduces the notion of an interpreter. The idea is that in order to have a representation, we need a system that interprets a certain state and acts accordingly (Godfrey-Smith, 2013). For instance, if we think about communication signals among animals, the claim that an interpreter is required for a state to qualify as a signal seems to be widely assumed in biology (Hauser, 1996). Similarly, this is a standard requirement in abstract models of signaling (Lewis, 1969; Skyrms, 2010) and I will argue it is also a common idea in neuroscience (see 4.1.2).[21]

But how does the existence of an interpreter help to determine representational content? A straightforward way of using the notion of an interpreter in a theory of representation is to claim that the function of a representation is to lead the interpreter to do something that is only successful when the represented state of affairs obtains. This is the key idea I suggest to develop in some detail.

More precisely, one of the first teleological views that introduced the notion of interpreter is the following (Papineau, 1984, p. 557; 1987, 1993):

Early Papineau

R represents S iff

1. R has a function in accordance with etiological function.
2. There is an interpreting system C with function F, such that the function of R is to contribute to C's performance of F.
3. S is the condition that must be mentioned in the most proximal Normal explanation of C's performance of F.[22]

I have not yet defined 'most proximal Normal explanation' stated in condition 3, but I hope that the main insight of Early Papineau can be intuitively grasped. In short, the idea is that the content of a representation is determined by the state of affairs that the interpreter

21 Here it might be useful to remember Dretske's (1988, p. 67) analogy: "Putting chilled alcohol in a glass cylinder does not generate a misrepresentation unless somebody calibrates the glass, hangs in on the wall and calls it a thermometer".

22 Papineau's early view (Papineau, 1987; 1993) was summarized in the following quote:

> 'The truth condition, for any belief, is that condition which guarantees that actions generated by that belief will fulfill its biological function of satisfying desires' (Papineau, 1993, p. 80).

There are two features that distinguish Early Papineau from the view expressed in this quote. First, Papineau was primarily interested in human cognition, so his approach to content is defined in terms of beliefs and desires (see chapter 5). However, here we are considering examples of cognitively unsophisticated organisms, like salamanders, frogs or beavers, which have very simple desires (usually called 'drives'), like the frog's desire for flies and the beaver's desire to avoid predators. So Early Papineau is a plausible extension of Papineau's insights to the representational capacities of simple organisms. The second feature that distinguishes Early Papineau and Papineau's historical view is that in the quote he appeals to the conditon that *guarantees* the satisfaction of desires (see also, McDonald and Papineau,2006). But this is surely too demanding. The presence of the represented state of affairs does not guarantee that the desire will be satisfied; there being a fly does not *guarantee* that the frog will have a meal. A more plausible proposal is that the presence of a fly is the crucial feature that explains how the frog usually enough satisfied his desire in the past (this is a rough gloss of condition 3). So, in this sense, I think Early Papineau is actually an improvement over Papineau's early view.

needs in order to perform its other functions successfully (Godfrey-Smith, 2013). For instance, the beaver's splash means *there is a predator around,* because beaver splashes have the function of leading other beavers to hide and avoid being eaten by a predator, and the normal explanation of how beavers performed this function must mention the fact that there was a predator around.

But, before assessing in detail whether Early Papineau is the right naturalistic account of content, let me spell out in detail what 'most proximal Normal explanation' means.

2.2.2.1 *Normal Conditions and Normal Explanations*

A *Normal explanation* (with a capital 'N', to mark that it is a technical notion-Millikan 1984) is an explanation of how a particular trait has historically performed its function. More precisely, 'a Normal explanation is a preponderant explanation for those historical cases where a proper function was performed' (Millikan, 1984, p. 34). The Normal explanation of how a heart performed its function must mention the fact that it was supplied with blood, it was connected to the rest of the body through the right vessels, and so on.

Indeed, there are many different features involved in the Normal explanation of how a device performs its functions. For instance, the Normal explanation of how a trait performs its functions typically mentions certain environmental facts. The Normal explanation of how my eyes perform their function must mention the fact that it is daylight and the fact that I am not wearing opaque glasses.

Two different kinds of Normal explanations must be distinguished: a *complete* and a *least detailed* Normal explanation. A *complete* Normal explanation mentions *all* facts that explain how a given trait has historically performed its function. For instance, the fact that gravity remained constant figures in the complete Normal explanation of how many consumer systems manage to perform their functions. A complete explanation contains a high number and variety of facts and many of them are irrelevant for content determination.

In contrast, the key insight of teleosemantics is usually expressed by appealing to the *least detailed* (or *most proximal*) Normal explanation (Millikan, 1984, ch.1). This is what Millikan (2002, p. 124) calls the 'Descriptive Generality Requirement'. The most proximal Normal explanation mentions those features that are really explanatory in the case at hand, so it does not appeal to standing circumstances (such as gravity or the absence of an Earthian explosion) or irrelevant details (such as which molecules were in fact involved in this process). For instance, the most proximal Normal explanation of how a heart circulates blood must appeal to the the oxygen supply, the presence of a closed circuit of blood vessels, and the regularity of electrical impulses sent to the heart (Millikan, 1984, p.33). A more detailed or less proximal explanation will tell us where the electrical impulses come from, and so on.

Now, according to traditional teleosemantic views, what determines the states a given representation is supposed to map onto is the state of affairs that must be mentioned in the *least detailed* or *most proximal* Normal explanation of how the interpreter performed its functions. The most proximal Normal explanation provides us with the key state that was specially needed by the interpreter in the situations where R was present. For instance, the circumstance that must be mentioned in the most proximal Normal explanation of how beavers performed

their hiding behavior is the presence of danger around. That gravity remained constant might figure in the *complete* Normal explanation, but surely not in the *most proximal* explanation.

The notion of Normal explanation helps us to define a second technical concept: Normal circumstances. Normal circumstances are those circumstances that must be mentioned in a Normal explanation. They are those circumstances that were present and figure in the Normal explanation of how a trait performed its function.

Crucially, notice that Normal circumstances need not be the most common ones. The Normal circumstances for sperm involve the presence of an ovum, since this is a condition that was present in those cases in which sperm performed its function, but obviously the presence of an ovum is not a statistically normal condition. Normality (with a capital 'N') has to do with those circumstances that explain why the trait nowadays exist. How often they occurred is irrelevant.

2.2.2.2 *Assessing* Early Papineau

Now we are in position to fully understand and evaluate Early Papineau. Early Papineau claims that representations have the etiological function of bringing consumers to do certain things, and the content of that representation is determined by the condition that must be mentioned in the most proximal Normal explanation of how the consumer performed this function. In the case of beavers, this condition seems to be the presence of a predator, because the most proximal Normal explanation of how the hiding behavior was performed must mention the presence of a predator. Thus, the content of the splash is something like *there is a predator around*.

But, is Early Papineau a satisfactory naturalistic theory of representation and content? Even if I think there are important insights in this approach that need to be preserved, I doubt Early Papineau can be right as it stands. Let me argue why I think this approach is utterly unsatisfactory:

effects First, Early Papineau suffers from the same problem as Crude Teleological Account, namely that of confusing the item that has the function with the state that is endowed with representational content. Prima facie, it seems that *traits* have functions and *states* have content. Or, more precisely, it seems that R is a contentful state in virtue of another trait having a certain function (see 3.2.6). Again, I do not think this is the most essential problem, but we will see that this is the central adjustment that might help to solve the rest of problems.

selection processes Secondly, Early Papineau will probably fail to explain the contents of human thoughts or, more generally, representations acquired by means of some learning process. Here is the reason: in order to satisfy 1, representations must have functions. Now, according to Etiological Function, a state can have a function only if it has been selected for in accordance with Selection For. But it is dubious that human thoughts or other kinds of representations undergo selection processes (this point will be argued in more detail in 3.2.6). So it is very unlikely that Early Papineau can account for the content of these complex representations.

Now, one could try to argue that thoughts actually undergo selection processes, by appealing to something like classical and operant condi-

tioning. Conditioning is a process that exhibits reproduction, random variation and differential replication, so it seems to fulfill the conditions required for Selection for (for a defense, Hull, et al. 2001; for a discussion, Artiga, 2010). Thus, it could be argued that this process generates functions in accordance with Etiological Function. However, notice that, even assuming that the kind of selection process that takes place in conditioning is a process that parallels natural selection, only if *all* sorts of learning were understood under the paradigm of classical and operant conditioning could one argue that learned representations possess certain functions. But the idea that all learning can be reduced to classical and operant conditioning has been largely discredited, specially in linguistics (Chomsky, 1959) and psychology (Fodor, 1975; Gallistell, 1990). Thus, a salient difficulty for Early Papineau is that there are plenty of representations, which have not undergone selection processes and hence lack functions, but nevertheless are endowed with content. Therefore, content cannot be analyzed in terms of the functions of representations.

compositionality Finally, the strategy of trying to account for the content of representations in terms of particular functions of those representations will probably be unable to explain one of the most important features of many representational systems, namely compositionality. Roughly, a representational system is compositional when the content of complex representations depends on the content of the simple representations it is composed of (and the way they are composed). Thought and language, for instance, are usually regarded as compositional. However, in general, functions of complexes do not decompose into the functions of their parts. The screwdriver has a function, and of course its parts have functions as well, but the function of the screwdriver is not composed of the functions of its parts.[23] So, contrary to Early Papineau, the content of a representation cannot depend on the function of the representation (for more on compositionality, see 6.1.1 and 6.5.1).

Therefore, it is very likely that Early Papineau cannot provide a general theory about representations and representational states. The crucial mistake (which, as we will see in 3.2.6, is also shared by some aspects of Millikan's theory of concepts, like the notion of 'derived function') is the assumption that in order to attribute content to representations we need to warrant *them* functions. Certainly, we all agree that functions are a crucial part of the teleological story, but attributing them to representations is a theoretical option- and one that yields the wrong results. Instead, I will argue that the crucial functions for content attribution are the ones had by the *systems* that produce states, rather than the functions had by the states themselves. Crucially, notice that the three objections to Early Papineau I pointed out are rooted in the assumption that representations themselves have functions. If we do not have to accept that representations have functions, we do not need to suppose that all of them undergo selection processes and we may be able to account for compositionality. Consequently,

23 At some places, Millikan seems to be suggesting that she disagrees: "Similarly, the sentence "It is raining" is a recurrent natural sign that the speaker believes it is raining. But it is not an intentional sign that the speaker believes it is raining. *Its memetic function, derived compositionally from the combined memetic functions of its significant parts*, is to produce beliefs that it is raining, not beliefs that speakers believe that it is raining' (Millikan, 2004, p.83. Stress added)

appealing to the function of states seems to be the wrong theoretical option.

There are, however, two ideas from Early Papineau that must be preserved. The first one is that we need to bring into the picture a system that interprets the representation. Otherwise it is hard to see how the notion of function can help to provide a naturalistic theory of representation and content. More precisely, the notion of an interpreter will be further developed in the more abstract framework of a sender-receiver structure. The goal of the next section is precisely to describe this sender-receiver model.

The second important idea, which I will develop a bit later (see 2.2.4), is that the content of a representation is determined by the condition that must be mentioned in the most proximal Normal explanation of how the interpreting system performed its functions. So the appeal to conditions for the proper performance of a function is a key concept in the proposal I will be defending in this thesis.

Thus, the task of the next sections is to spell out the sender-receiver framework and explain the conditions for content determination. This discussion will eventually lead to the first version of the teleosemantic account I would like to defend.

2.2.3 *Sender-Receiver*

As we just saw, apart from the notion of function, there is a second important concept that needs to be taken into account in any plausible teleological account of content: sender-receiver structure. This notion is familiar to several theories of information, like Game Theory and Signal Detection Theory (Lewis, 1969; Skyrms, 1996; 2010, p. 7; Godfrey-Smith 2006). The sender-receiver framework has also been used by some teleosemanticists (Millikan, 1984, 1993; Godfrey-Smith, 1996, 2013), although they provide different reasons for using this framework. According to the sender-receiver model, representations are states that stand between a system that sends them and a system that receives them. To a first approximation, a sender (also called *producer* system) can be defined as a system that takes some external cues as inputs and generates the representation as output, whereas the receiver (*interpreter* or *consumer* system) is the system that takes the representation as input and generates certain activity as output.[24]

Even though people working on abstract models talk as if a sender-receiver structure was all that is required for a mechanism to qualify as a signaling system (Skyrms, 2010), this structure fails to fully specify a set of necessary and sufficient conditions for a representational system to arise. A thrown ball which hits another ball seems to instantiate the schema I just offered: the first ball takes an input (its throwing it) and yields an output (the hitting) and the second ball takes the hitting as input and produces movement as output.[25] But, obviously, we do

24 I think that the literature on teleosemantics has failed to appreciate the crucial importance of the sender-receiver structure (see, for instance, Papineau, 1993; Neander, 1995; Stegmann, 2009; Cao, 2012). Notable exceptions are Godfrey-Smith (1996, 2013) and Millikan (1984), even though this aspect of their views has not received enough attention.

25 One could not reject this example by simply replying that balls are not systems. In teleosemantics, it is important not to interpret 'system' in any loaded sense. A magnetosome in certain bacteria or a set of neurons can be a system in the minimal sense intended here. Arguably, it should be understood in roughly the same way Machamer et al. (2000) use the term 'mechanism'.

not want to say that a *ball's hitting another ball is* a representation of anything.

So we need to set up several conditions that must be met for a structure to qualify as instantiating a sender-receiver model. Here is where models from signal detection theory should be complemented with the notion of etiological function that we presented above. The idea is that a structure instantiates the appropriate sender-receiver model if, and only if, there are two systems with certain etiological functions (to be specified below).

Indeed, notice that this proposal is supported by the previous discussion of Crude Teleological Theory and Early Papineau. First, it fits with the intuition that the notion of function is more naturally applied to traits or systems, rather than states. According to the framework I am suggesting, the kind of systems that are endowed with functions are producer and consumer systems. Secondly, we saw that the problems of Early Papineau derived from assuming that particular representations are endowed with functions. In the picture I am offering, representations need not be endowed with any kind of function; only systems do.

The idea, then, is that by adding the notion of function to the sender-receiver structure, we kill two birds with one stone. On the one hand, and connecting with the result of previous sections, applying the notion of function to the *systems* that produce and consume the representation will enable us to overcome the objections to Crude Teleological Theory and Early Papineau and keep the intuition that functions are most naturally attributed to systems. On the other, we will have a principled way of individuating the relevant kind of producer and consumer systems, so that we can avoid saying that balls can instantiate a sender-receiver structure (balls do not have etiological functions). So there are good reasons for bringing in the concept of (functional) sender-receiver structure to our teleosemantic theory.

Furthermore, let me stress that the appeal to sender-receiver structures is well-entrenched on several sciences that study signaling, communication or information. For instance, ethological studies on animal communication are pervaded with ideas like the following:

> Thus, information is a feature of interaction (...). between sender and receiver. Signals carry certain kinds of informational content, which can be manipulated by the sender and differentially acted upon by the perceiver (Hauser, 1996, p.6)

> Instances of communication involve not only a physical *signal* such as light, sound or odor, but also a *sender* and a *receiver*. (...) The physical properties of a signal and the perceiver's perceptual sensitivity should be matched to each other and to the transmission properties of the environment. (...). The meaning of the signal, on the other hand, is inferred from the behavior of the receiver, so it may vary with the receiver's characteristics. (Shettleworth, 2010, p. 512)

Similarly, I will show in 4.1.3 that, appearances notwithstanding, neuroscience also adopts a sender-receiver paradigm. If we add the Signal Detection Theory mentioned earlier, we can reasonably conclude that there is some scientific evidence for holding that a sender-receiver structure constitutes a central kind in the context of signaling systems.

In accordance with these ideas, let me suggest a first definition of sender-receiver structure:

First Sender-Receiver

> Any two systems P and C[26] configure a sender-receiver structure if, and only if:
>
> 1. P and C have functions in accordance with Etiological Function
> 2. P and C have coevolved in such a way that a Normal condition for the proper performance of each is the presence and proper functioning of the other.[27]
> 3. P has two functions:[28]
>   a) The *non-relational* function of helping C to perform its functions.
>   b) The *relational* function of producing state R when another state of affairs S obtains.
> 4. The function of C is to produce an effect E. The least detailed and most comprehensive Normal explanation for C's performance of E involves S.

Let me briefly explain and justify each of these conditions.

(1) I hope the need for 1 is sufficiently clear from what I said earlier. The appeal to functions allows us to exclude mechanisms that merely instantiate a certain causal role, like the ball's example. Furthermore, it has all the advantages of teleological theories we pointed out at he beginning of the chapter without falling prey to the problems of Early Papineau.

(2) Condition 2 includes two claims that could be distinguished. On the one hand, the idea that sender and receiver must have coevolved. On the other, the more demanding claim that sender and receiver must have coevolved *as cooperating devices* (more precisely, such that a Normal condition for the proper performance of each is the presence and proper functioning of the other). The latter entails the former, but we can present support for each claim separately.

The first (rather uninteresting) reason for holding that a sender and receiver must have coevolved is that coevolution is required in order to avoid the following counterexample: one could artificially take a certain producer system and tie it up with a different consumer system; without condition 2, that would immediately generate a new sender-receiver structure, and hence the representation would change its meaning. But, intuitively, this is not the kind of structure that generates representations. For instance, we want to resist the claim that removing the frog's visual system and inserting it within (say) a snake, would automatically change the content of the mental representation. This sort of examples can be avoided if we add the constraint that both producer and consumer systems must have coevolved.[29]

26 P and C refer to types.

27 In other words, P and C are cooperating devices.

28 Notice that neither of the two functions is superfluous because P can comply with one function without fulfilling the other. Both are effects of P that explain its maintenance.

29 One might worry that there are some cases where there has not been coevolution and, nevertheless, we want it to qualify as a sender-receiver structure. For instance, we could

Condition 2, however, not only appeals to coevolution but also to cooperation. The intuitive idea behind this claim (which is also supported by abstract models of signaling systems such as Lewis, 1969) is that, on the one hand, senders acquire the capacity of producing signals only if they are benefited from the receiver's performance of its functions; otherwise, they would stop producing signs (Millikan, 2004, 2005). In turn, receivers benefit from the activity of senders; otherwise, they would stop attending to signs (Shettleworth, 2010, p. 513). That shows that a Normal condition for the proper performance of each is the presence and proper functioning of the other; in other words, sender and receiver must have evolved as cooperating systems (which, of course, entails that they have coevolved). These ideas will be developed in more detail in 3.3.3.

Notice that condition 2 holds true even if there is only partial common interest between sender and receiver (Skyrms, 2010, Godfrey-Smith, personal communication). It suffices if the Normal condition for the performance of one's function is that the fact that that the other system also performs its function. The only scenario that is ruled out is the existence of a signaling systems in which there is no common interest between the parts (see Skyrms, 2010, p. 77-8).[30]

(3) Condition 3 involves a distinction between relational and non-relational functions that needs to be explained.[31] A device has a relational function if its function is to do or to produce something that bears certain relation to something else (Millikan, 1984, p. 39). The paradigmatic case is the chameleon's pigment-rearranging device. It is widely known that chameleons change the color of their skin so as to match the surface they are sitting on. Arguably, the function of this pigment-rearranging device is to produce a color that is supposed to correspond with the color of the surface.[32] Notice that this function (the effect that explains why the device has recently been selected for) is to produce a certain state *when* another state of affairs obtains. This is the key property that distinguishes them from non-relational functions, which lack this relational aspect. Kidneys, for instance, have the non-relational function of filtering wastes from blood.[33]

---

implement an artificial eye to a snake, and we might want this cyborg to instantiate a sender-receiver model; after all, it seems it has full-blown visual representations.
The classical response to this sort of problem is to point out that this eye would be an artifact, which would have probably been designed by someone with a certain purpose. So it might be that the intentional properties of the artifact derive from the intentional properties of the designer. As I stated in the introduction (see 2.1.2), how artifacts acquire their functions is a very complex issue I will not attempt to address in this dissertation. In any case, as I argued in chapter 1, I think that any satisfactory answer to the problem of semantic properties of artifacts requires a naturalistic account of mental (semantic) facts, which is what I am trying to provide here. If the worry is extended to undesigned traits or organisms, see the discussion on Swampman in 3.3.4.

30 Thanks to Manolo Martinez for pressing me on that point.

31 Of course, this distinction is originally motivated by Millikan's notion of direct and relational function. On Millikan's terminology, relational functions are a special kind of direct functions ('Some direct proper functions are relational', Millikan, 1984, p.49). Unfortunately, she does not coin any name for those direct functions that are not relational. This is the reason I introduce the name 'non-relational function'. Nothing I say here is supposed to contradict Millikan's insights.

32 Some recent studies suggest that, as a matter of fact, this ability might have a signaling function among conspecifics, rather than a camouflaging function (Stuart-Fox and Moussalli, 2008). Nevertheless, the function that was traditionally attributed to the chameleon's skin color is actually exemplified in other species (e.g. certain cephalopods), in which a change in color skin primarily has a camouflaging function.

33 As I will argue in the next chapter (section 3.2.2), the chameleon pigment-rearranging device has not only a relational function, but what I will call an 'open relational function',

Now, condition 3 has two parts. First, 3a claims that producer systems have the (non-relational) function of helping C to perform its functions. The expression 'helping' should not raise naturalistic qualms; the same idea can be expressed by saying that a Normal explanation for the proper performance of the producer's function is the fact that the consumer performs its own functions. That claim naturally follows from the assumption that they are cooperating systems. If a Normal condition for the proper performance of each system is the presence and proper functioning of the other and given that the main effect of the producer has a direct causal influence on the receiver, then it seems reasonable to suppose that one of the functions of the producer system is to help the consumer to perform its own functions. So 3a is, I hope, clearly justified.

A bit more controversial, though, is 3b, the idea that a (relational) function of sender is to produce a state R that is supposed to correlate with another state S. The idea here is that the way the producer system achieves the (non-relational) function of helping the consumer system is by producing a state (the representation) when another state of affairs obtains (the representatum).[34] This is a function of the producer because this is an effect that helps to explain why it has been selected for. This, however, is a contentious claim because some teleosemanticists deny that the function of a system can be to produce a state that correlates with another state (Millikan, 2004; Papineau, 2003; for an exception, Dretske, 1988). In 2.2.3.1 I will address some worries this claim may raise; before that, let me briefly justify the last condition of Sender-receiver.

(4) One of the key insights of teleosemantics is that the content of a representation is determined by the the state of affairs that Normally is required by the consumer system in order to perform its functions. So what R is supposed to map onto utterly depends on the needs of the consumer system. As I claimed earlier, one of the features of Early Papineau that I think must be maintained is the idea that content is determined by the condition that must be mentioned in the relevant *Normal explanation* of how the consumer system historically performed its functions. Condition 4 directly derives from condition 3 in Early Papineau.

Now, let us focus on the idea that the relevant Normal explanation is the *least detailed* and *most comprehensive*. Remember that there are many different kinds of Normal explanations: the complete explanation, the most proximal explanation and so on. In traditional teleosemantic accounts such as Millikan's, the third condition has been exclusively formulated in terms of the *least detailed* Normal explanation. However, they also assume that this explanation should in some sense be *satisfactory*. The relevant Normal explanation is the one that is least detailed and, nevertheless, provides a satisfying explanation. More precisely, I think the main idea should be better spelled out in the following terms: S (the represented state of affairs) is the feature that must figure in the

since it can produce colors that had never existed before. For the moment, though, it is enough if we keep in mind the notion of relational function and its contrast with non-relational functions.

34 From the way I defined a sender-receiver structure, it seems to follow that senders can only produce one kind of representation, namely R. But, one might argue, many representational systems can surely produce new kinds of representations. These and other complexities will be developed in the next chapter, where I will present a more complete and nuanced definition of the sender-receiver framework.

least detailed and, nevertheless, most comprehensive Normal explanation. The relevant feature, then, is the one that gets the best result from the trade-off between being the least detailed and, nevertheless, the most explanatory feature of the past success of the behavior of the consumer system (for a more detailed analysis, see 2.3.3). I think this second aspect of the Normal explanation is crucial and has often been obscured by traditional theories. So the particular formulation I put forward is original and, nevertheless, clearly inspired by traditional accounts. This idea will develop in more detail in several places of this dissertation.

Now, someone might object that the notion of 'explanation' does not seem to be a naturalistically acceptable concept; if the teleosemantic recipe for attributing content utterly depends on the notion of explanation, one might think we are not reducing content to non-intentional notions. However, these worries are ungrounded. One could express the same idea in a different way: the represented state of affairs is the state that primarily *causes* the fitness enhancing profile of the representational system (Martinez, 2010). In other words, what primarily causes the beaver's hiding behavior to be fitness-enhancing is the presence of a predator, so this is the condition that must be mentioned in the least detailed and most comprehensive Normal explanation of how beavers performed their hiding behaviors. This is precisely what justifies the appeal to the least detailed an most comprehensive Normal explanation: by selecting this kind of explanation, we are trying to pick up the feature that is most causally relevant in the success of the consumer system. So, at the very end, the appeal to the least detailed and most comprehensive explanation is not arbitrary at all, but grounded on a causal relation. Of course, the notion of causation is itself extremely controversial; nevertheless, I hope it is clear that appealing to Normal *explanations* does not necessarily compromise our naturalistic project.

Let me now turn to one of the most contentious claims of Sender-receiver, namely the idea stated in 3b that one of the (relational) functions of the producer system is to produce a state R that is supposed to correlate[35] with another state S.

2.2.3.1 *The Functions of Producers*

A possible objection against Sender-receiver is the following: condition 3b claims that the (relational) function of a system is to produce a state R when another state S obtains. That is, the function is to produce a state that correlates with another state. Now, many people have objected that 'correlating with food' or 'correlating with black moving things' cannot qualify as functions (Millikan, 1993; Burge, 2010, Ch. 10; Papineau, 2003). Paradigmatic effects of systems are *fleeing from a predator, catching flies, getting food*, etc... but correlating with such and such does not seem to be among the system's effects. And, of course, if correlating is not a system's effect, it cannot be its function either.

In what follows I will try to present and argue against different arguments one might hold in order to resist the idea that systems can have the function of producing R when S obtains. The goal is to defend the condition 3b.

35 For reasons that will become clear in the next chapter, the notion of 'correlation' is not the right one for describing the relational function of producers (see 3.3.2.1). Nevertheless, it is probably a ladder that must be used at this stage, and only thrown away after presenting the whole teleosemantic account in the next chapter.

FUNCTIONS OF SYSTEMS First of all, it is important to be aware of what are we attributing a function to. If we think that we were trying to find out the function of a *representation* (a view I will extensively argue against in 3.2.6), then certainly, detecting does not seem to be an effect of a representation. By definition, representations have content, so the idea that an *effect* of a representation is to correlate with its content seems misguided. However, I hope I have been able to show that teleosemantics should not be concerned with the function of representations, but the function of mechanisms producing representations. And once we focus on the function of systems, the idea that a function of a device is to produce a state R when another state S obtains is completely transparent.

METAPHYSICAL WORRY We can interpret this concern as raising a metaphysical worry: can 'state R correlates with S' be considered an effect? More generally, can a correlation be considered an effect at all? If this is the kind of concern, I think a brief reasoning can provide a satisfactory answer.

Functions are effects and effects are states of affairs (or facts, as I am using both terms interchangeably). On one standard view of what states of affairs are, they are composed of objects that instantiate certain properties (Armstrong, 1997, 2004). The details about the metaphysical relation between objects and properties need not concern us here. If something like this view is the right way to think about state of affairs and assuming that effects are states of affairs, then I do not see why 'state R correlates with S' cannot be considered an effect, and hence a function. There fact that R and S obtain at the same time seems to be a state of affairs as any other.

The same point can be put in a different way. The state R (the mental state that we regard as the representation) is in itself a state of affairs (e. g. activation of a neuronal network); so, strictly speaking, producing a mental state R could already be the function of the producer system, because it is an effect. But if producing state R is an effect (and hence, could qualify as a function), then producing an effect R when S seems a perfectly well-formed state of affairs. And if it is a state of affairs, there is no metaphysical reason why it cannot be considered an effect, and hence a function.

AN EFFECT OF P? A related worry is that even if it is granted that 'R correlating with S' qualifies as an effect, one might argue it is not an effect *that is caused* by the system. The system, one might argue, only produces R, not the correlation of R with S. So R correlating with S cannot be a function *of the system*.[36]

That objection seems to assume a wrong conception of what it is to be an effect of a system. Surely, a state F can be an effect of S even if many features of F have not been produced by S, or even if S only causes F if many other conditions also hold. For instance, it is widely accepted that the bacterium *Mycobacterium* causes tuberculosis. So tuberculosis is an effect of an agent's body being infected by *Mycobacterium.* However, *Mycobacterium* is a necessary but not sufficient cause of tuberculosis, since some people carry this bacterium yet remain entirely asymptomatic. Complementary factors are required, such as genetic susceptibility, poor nutrition, familial exposure or immunosuppression (Gertsman, 2003,

36 Thanks to David Pineda for raising this objection.

p. 42). Similarly, the fact that a producer system creates R seems to be a necessary condition of R's correlation with S, but other conditions must also hold. If tuberculosis is an effect of *Mycobacterium* being in an organism, it seems *R correlating with S* can also be an effect of the representational system. Therefore, I think producing R could suffice for *R's correlating with S* to be an effect of the system

Indeed, in this response I am probably granting too much to the objector. For it seems that the system (type) is partially responsible of the correlation itself. Here is the reason: it follows from the definition provided above that if there had been no correlation between R and S, there would have been no representational system (or, at least, a very different representational system would have evolved). So, in fact, the Reproductively Established Family to which a given representational system belongs is at least partially responsible of the correlation between the representation and the representatum. So, indeed, R's correlating with S can be a function of the system not only because the system produces R (which is a necessary condition for the correlation to occur) but also because the system has been causally relevant in the maintenance of this correlation. Therefore, I think we can reject this criticism.

CORRELATING AND FITNESS Fourthly, one might argue that it is not obvious why the fact that R correlates with S can increase fitness in any relevant sense (Burge, 2010; Papineau, personal communication). Effects like *ingesting a fly* or *escaping from a predator* obviously increase the fitness of organisms, so they can explain why the mechanisms that produce them have certain etiological functions. But (the objection runs), R correlating with a state S does not look like an effect that can contribute to an explanation of why the producer system exists. If that is right, the fact that R correlates with S cannot said to be a function.

Burge (2010), for instance, has recently argued along these lines. He claims that R correlating with S is only fitness-enhancing because it causes certain behaviors on the consumer system like feeding or mating. Consequently, only the latter should qualify as functions in a proper sense:

> I do doubt that biological functions, as ordinarily understood, ever reside strictly on detection by itself, or in mere correlation with distal conditions. (...)
>
> Detection is, however, not in itself a biological function, as 'biological function' is standardly understood. Detection failure is not in itself failure of biological function. It is the contribution to response, and ultimately to fitness, not the detection per se, that is biologically functional. Detectors were selected, not because they were accurate in detecting a condition, but because they tended to contribute fit responses, including fit behavior, with respect to that condition. (Burge, 2010, p. 302)

Now, I think this argument derives from a misunderstanding of biological explanations. An effect can contribute to the selection of a trait, even if the advantage it confers is extremely small. The function of eyebrows, for instance, is to protect the eyes from sweat, water and debris. Of course, one might argue they also have the (more distal) function of enabling vision (by protecting the eyes), but this claim by no means

contradicts the more proximal function. Similarly, finger nails are only useful because they protect our fingertips from injuries, but nevertheless they are functional traits. Filtering wastes from blood is only useful because it contributes to the general homeostasis of the organism and, nonetheless, this contribution is said to be a function of kidneys. In general, the claim that functional traits are only those that directly bear on feeding or mating behavior is surely mistaken. Any effect can be selected for, if there has been a process that satisfies SELECTION FOR. If only effects like *ingesting food* could ground an attribution of functions, very little of our traits would be functional. And once we adopt this more sensible understanding of natural selection, the fact that a producer system produces a sign when another state obtains seems to provide a significant advantage over organisms lacking it. Therefore, R correlating with S can not be dismissed as being an irrelevant effect.

EXAMPLE To conclude this defense of correlations being functions, let me present an example where it seems perfectly obvious that the (relational) function of a device is to correlate with something else. Cuttlefish are usually labeled 'the chameleon of the sea' due to their remarkable capacity to change their skin color in order to camouflage from predators. They have up to 200 of specialized pigment cells per square millimeter, which allow them to produce the color of the surface the cuttlefish is laying on. It seems quite plausible that the function of the pigment-rearranging device is to produce a color that matches the color of the surface the cuttlefish is sitting on. Even if the cuttlefish's color is not a representation (because there is no adequate consumer system), the function of the pigment-rearranging device is exactly the same kind of function as the one had by (representational) producer systems. In the case of colors as well as representations, the function is to produce *something when something else is the case*. So, if one admits that this can be the function of the cuttlefish's pigment-rearranging device (and I think this function is hard to deny), there seems to be no reason for rejecting the idea that this is the function of standard producer systems in sender-receiver structures.

2.2.4 *Representations and content*

The sender-receiver structure set up above is the key element that enables us to define the rest of notions in our basic teleosemantic framework. On the one hand, a representation is a state that stands between a producer and a consumer system. In turn, both systems must have certain etiological functions (in accordance with ETIOLOGICAL FUNCTION), that is, the producer and consumer system must have been selected for because they have historically performed certain tasks that are somehow useful for the organism. In particular, the function of the sender is to produce a state (what we call the *representation*) when another state of affairs obtains (the *representatum*). In other words, the function of the system that generates representations is to produce a state that covaries with certain environmental circumstance which has some interest for the organism (resources, predator,...). On the other hand, the function of the consumer system, which receives and interprets these representations, is to perform certain activities that (in Normal conditions) are successful because the represented state of affairs obtains. The most proximal and most comprehensive Normal

explanation of how the consumer system performs its functions must mention the state that the representation is supposed to correlate with. Hence, while the function of the sender is to produce a representation that corresponds with certain world affairs, the receiver has the function of performing other activities, which Normally are successfully carried out because there is a match between the representation and the represented affairs.

Hence, on this framework we can define with certain precision what representations are:

First Representation R is a representation iff R is a state produced by a sender P, which satisfies First Sender-receiver.

Hence, whether a state is a representation or not depends on the existence of an adequate sender-receiver structure, where sender and receiver are endowed with the right etiological functions. Accordingly, the content of a representation (what a representation is supposed to map onto) is determined by the functions of the producer and the consumer system. More precisely:

First Content

R represents S iff there are two systems P and C such that:

1. P and C configure a sender-receiver structure, in accordance with First Sender-Receiver.
2. R is a representation, in accordance with First Representation.
3. The most proximal and most comprehensive Normal explanation for C's performance of its functions when R obtains involves S.[37]

Let me illustrate First Representation and First Content with an example. Red-backed salamanders (*Plethodon cinereus*) have an internal mechanism that is sensitive to certain odors which are usually produced by predators, specially by eastern garter snakes (*Thamnophis sirtalis*) (Sullivan et al. 2001). When this cue is present in the water surrounding the salamander, it significantly reduces its foraging behavior (Sullivan et al. 2001). Now, there is an internal state (a signal) that is produced by the mechanism that is sensitive to the odor, which in turn is interpreted by another mechanism that reacts by reducing foraging activity. In this case, the definition provided above predicts that the internal sign is a representation and its content is something like *there is an eastern garter snake around*, since this is the state of affairs that must be mentioned in the least detailed and most comprehensive Normal explanation of how the interpreter performed its functions. The least detailed explanation why Normally Red-backed salamanders reduce foraging behavior must mention the presence of eastern garter snakes. In the evolutionary past, this reduction of activity was mostly helpful when there was a snake around; when there was no threat, the behavior was a lost chance for mating or ingesting food.

Now, the teleosemantic framework is constituted by jointly endorsing First Sender-Receiver, First Representation and First Content.

37 Needless to say, First Content will have to be refined in several ways, specially in order to account for systems that can represent states of affairs that have not existed before. This is one of the tasks I carry over in the next chapter. Nonetheless, I think First Representation and First Content suffice for responding some of the most common objections to teleosemantics.

Hence, I will use the expression 'First Teleosemantics' in order to refer the theory constituted by these three definitions.

Crucially, notice that First Teleosemantics provides a naturalistic account of content. To say that a state R refers to S just means that the complex relation stated in First Teleosemantics holds between R and S. In other words, a particular instance of the structure described in First Teleosemantics is the truthmaker for the claim that R means S. And notice that in the *explananda* we have not used any intentionally loaded notion: Selection for, etiological function, sender-receiver, representation, First Representation and First Content have been defined in naturalistically acceptable terms. So, as far as naturalism is concerned, First Teleosemantics seems to be a satisfactory naturalistic account of content.

2.2.4.1 *Semantic and Metasemantic Theories*

Before moving to some objections, let me return to one of the questions that we discussed in the previous chapter. We saw that most current naturalistic theories are only semantic theories of content (i.e. they purport to explain in virtue of what process a given state means A rather than B), but not metasemantic theories (i.e. they do not try to explain why some states represent something at all). We are now in position to explain why teleosemantics provides a *metasemantic* theory besides a semantic one. First, there is a reason which is, as it were, *internal* to the theory: in order to introduce the notion of function in a satisfactory way, we have to appeal to the systems that produce representations (this was the main lesson from Early Papineau's failure). As a result, teleosemantics needs to define representational systems (sender-receiver structures), and thus it has to explain what representations are. More generally, we saw that any plausible teleological theory has to appeal to representational systems, and hence it must provide a metasemantic theory of representation.

Secondly, as we saw in chapter 1, there is an *external* motivation that derives from our naturalistic goal: for a *naturalistic* theory to fully explain what content is, it has to explain what representations are, and hence it has to provide a metasemantic account. Any purely semantic approach that only tells us in virtue of what process a representation refers to A rather than B, but remains entirely silent on what makes a state a representation at all, would only partially clarify the main perplexity caused by intentional properties. Providing a semantic theory only goes halfway towards a full naturalization of content. A metasemantic account is required in any complete and satisfactory naturalistic account of representation and content.

The two reasons push in the same direction: teleosemantics has to be able to provide a satisfactory semantic and metasemantic account of representation and content.[38] Obviously, that carries with it certain

38 While most philosophers interpret the theory as a metasemantic one (e.g. Matthen, 2006, p. 149), Millikan seems to have changed her view in that respect. In Millikan (1993, p.123) she seemed to be agreeing with me, but consider the following quotes from some recent work:

> Accordingly, naturalist theories of the content of mental representation are often divided into, say, picture theories, causal or covariation theories, information theories, functionalist or causal role theorists and teleological theories, as though all these divisions all fell on the same plane. That is a fairly serious mistake, for what teleological theories have in common is not any view about the nature of representational content. "Teleosemantics",

explanatory advantages over the accounts we discussed in chapter 1, but it also brings with it new difficulties. The rest of the chapter is devoted to analyze whether teleosemantics can meet these challenges.

2.3 OBJECTIONS

In chapter 1, I distinguished four problems that any naturalistic account of content should address. Of course, they are not the only objections (chapter 3 is precisely devoted to reply to several other difficulties), but a theory that is unable to solve any of these four problems is a non-starter. Thus, it is time to assess whether First Teleosemantics can deal with them.

2.3.1 *Misrepresentation and Normativity*

The two objections that are more easily dealt with by First Teleosemantics are the Error and the Normativity Problem. Let us briefly show how First Teleosemantics overcomes these objections.

Consider the Error problem as was formulated in 1.2.2.2:

(Error Problem) A semantic theory suffers form the Error problem if it does not allow for cases of misrepresentation.

First Teleosemantics can easily account for cases of misrepresentation. R misrepresents when the consumer system has the function of producing R when S obtains, R is produced but S does not obtain. Furthermore, notice that, in contrast to Strong indication, R can represent S even if R is false most of the time. Since a trait can have a function that it only rarely performs, a producer system can have the function of producing R when S obtains, even if most of the time R is produced when S does not obtain (that is, even if R is false most of the time). Moreover, in contrast to weak indication and relative indication, prima facie content seems to be precise and adequate, since it is determined by the state of affairs that the consumer system has historically required in order to perform its function in a Normal way (but see below). So, First Teleosemantics is able to account for misrepresentation without falling into the obvious drawbacks of Causal Theories.

On the other hand, since First Teleosemantics is intended as a metasemantic theory, it also has to address the Normativity problem:

(Normativity Problem) A metasemantic theory suffers from the normativity problem if it cannot account for the normative difference between cases of successful representation and cases of misrepresentation.

---

as it is sometimes called, is a theory only of how representations can be false or mistaken, which is a different thing entirely. Intentionality, if understood as the property of "offness" or "aboutness", is not explained by a teleological theory. Natural signs are signs of things and represent facts about things, but they cannot be false. To explain the possibility of falseness, then, cannot be the same as to explain ofness or aboutness.(Millikan, 2004, p. 63)

The teleologist needs a base theory of the representing relation [picture theories, causal theories or the like] on which to build his description of intentional representation. (Millikan, 2004, p.71).

First Teleosemantics seems to be specially well equipped for dealing with Normativity. Cases of misrepresentation are wrong, because they are cases in which the representational system fails to fulfill its function. In the same way that kidneys *are supposed to* filter wastes from blood, and if they do not filter wastes we say that they *mal*function, *dys*function or just function in the *wrong* way, cases of misrepresentation involve the same kind of normativity. Similarly, we can explain why representing truly is right by appealing to the past situations in which true representations led to success. Hence, there is a close connection between truth and success and falsity and failure.[39]

Therefore, it seems First Teleosemantics can readily overcome the Error and the Normativity Problem. Unfortunately, things are more complex with the Indeterminacy and the Adequacy problem.

2.3.2 *Indeterminacy and Adequacy*

Does First Teleosemantics solve the indeterminacy and adequacy problems laid down in the previous chapter? Many people think it does not. The relevance of this objection in the context of teleosemantic theories is so great that I will spend the remainder of this chapter trying to show how this objection can be met within the framework set up above.

In order to develop the difficulty in some detail and assess the different views, it might be worth explaining in some detail the case of leopard frogs, which has dominated the literature on this problem since the 80s.

hunting in leopard frogs The aspect of leopard frogs (*Rana Pipiens*) that has centered the vast literature on the indeterminacy problem is its hunting mechanism, which has been extensively studied by ethologists since the 50s. There are two biological mechanisms involved in the hunting behavior of leopard frogs: the visual system and the tongue-snapping mechanism.

On the one hand, the frog's visual capacity is far less accurate than ours. They can only distinguish black shadows moving at a certain distance, so they are unable to differentiate bees, pellets, flies or any other small object that casts a black shadow and moves at a certain velocity (for a detailed description of the anuran visual system, see 4.2.2). Nonetheless, they have evolved a quite successful hunting mechanism: whenever they detect a black moving thing passing in front of them at a certain distance and velocity, they throw their tongue out and catch whatever they find there (Lettvin *et al.*, 1959). Obviously, due to their poor visual mechanism, many things can elicit this hunting mechanism, but the key point is that in the environment where frogs evolved, usually enough this black moving things were flies.[40]

39 One might worry that this close connection between truth and success entails that (according to teleosemantics) attributions of *true* representations can not provide substantive explanations of *successful* behavior (because true representations are partially explained by appealing to successful behaviors). This objection will be addressed in 3.3.2.

40 As it is common in the literature on this topic, I am going to assume throughout this chapter that frogs only prey on flies. Even though it is empirically false (Neander, 2006), this simplification is going to be very helpful in order to keep the example as simple as possible. In the next chapter I'm going to drop this assumption and show how my solution can deal with more realistic scenarios. In chapter 4, I will provide a detailed description of the visual system of the European toad (*bufo bufo*).

Let me first present the problem suggested by this example in an intuitive form. If we focus on the hunting mechanism of frogs, the problem is that it seems that First Teleosemantics warrants many different content attributions to the frog. For example, someone might claim that First Teleosemantics warrants to the frog's mental state the content *there is a black moving shadow.* In the environment where frogs evolved, most black moving shadows were flies[41], so the correlation between the frog's mental state and there being a black moving shadow can explain why the device was selected for by natural selection. Similarly, the fact that the mental state correlated with flies can also explain why this mechanism was advantageous. So, even if it follows from First Teleosemantics that frogs represent *there is a fly around,* it also seems to follow that the content of the frog's mental state is *there is a black moving shadow*:

> Notice that, just as there is a teleological explanation of why frogs should have fly detectors- (...) so too there is a teleological explanation of why frogs should have a little-ambient-black-thing-detector (...). The explanation is that in the environment in which this mechanism Normally operates all (or most, or anyhow enough) of the little ambient black dots are flies. So, in this environment, what ambient-black dots detectors Normally detect (de re, as it were) is just what fly detectors detect (de dicto, as it were); viz. flies. (...) Correspondingly both ways of describing the intentional objects of the snaps satisfy what Millikan (1986) apparently takes to be the crucial condition on content ascription. (Fodor, 1990, p. 72)

More generally, the problem is that there are many possible states of affairs whose representation could be advantageous for frogs. The content teleosemantics yields is not determinate enough. Contents that could explain why the representational system was selected by natural selection are, among others: *there is a black moving shadow, there is a fly, there is a bug, there is a black dot in the retina....* The objection then, is that we usually think (and ethologists seem to suggest) that the content of the frog's mental state is more determined than the one provided by teleosemantics. This is the key point of the Indeterminacy Problem, as formulated in 1.2.2.4:

(Indeterminacy Problem) A theory suffers from the indeterminacy problem if it warrants multiple content attributions in cases where science and common sense warrant a single content.

Of course, if the representational content attributed by teleosemantics is so indeterminate, it will also fall prey to the Adequacy Problem. Consequently, unless Teleosemantics provides a principled way of picking out one among these different content attributions, the theory will utterly be unsatisfactory.

Interestingly enough, one of the most surprising facts about the vast literature on this objection is that it has shown that, even if most teleosemanticists do not accept that content is so indeterminate, there is no agreement as to what the right content is supposed to be. That is: different teleological theories of representation attribute different

41 Notice that it is not even necessary that *most* black moving shadows be flies. It is only required that *enough* of them be flies.

contents to the frog's mental state, and all of them think they agree with science and common sense. For instance, one important strand championed by Neander (1995, 2006) claims that the frog's mental state represents something like *there is a black moving thing* (for a discussion, see 3.3.1). On the contrary, Millikan (1984, 1993) has argued that the content should be something like *there is a frog food.* Still other people have defended that content is *there is a fly*, *nutritious stuff*,... and so on (Sterelny, 1990; Price, 1998, 2001).

Now, I think all current teleosemantic theories fall short of providing a satisfactory solution to this problem. This point will be argued when alternative teleosemantic theories are presented (see 3.1.2, 3.3.1, 5.2.3.1). Indeed, I think that Millikan has also failed to provide an adequate reply, even if her theory has the resources for dealing with it.

In this next section I will present an reply to this pivotal objection, relying on First Teleosemantics.

2.3.3 *Two versions of the Indeterminacy Problem*

As we said in the introduction, the Indeterminacy problem has been presented under many different names and classifications. I think the basic idea can be cashed out in two different ways, which I call the 'vertical' and the 'horizontal' problem.[42] This is important because each reading gives rise to a different version of the problem and requires a different answer. The horizontal problem has to do with the indeterminacy among states of affairs that lie at different levels of distality, while the vertical problem has to do with the different properties involved in a state of affairs.

Let us present the two problems and consider what teleosemanticists should say about them.

2.3.3.1 *The horizontal problem*

Basically, the horizontal problem consists in the fact that there are many states of affairs that correlate with the mental state R and whose correlation can explain why the representational system was selected for. Consequently, there are many states of affairs that could be said to be the content of the representation. Consider again leopard frogs. Since usually enough black shadows are produced by flies, representing the presence of a black moving shadow could explain why the representational mechanism provided an advantage to organisms having it and hence why organisms with such a mechanism were selected for by natural selection.

Of course, moving black shadows are not the only problematic candidates. Think about the black dots in the retina that are usually caused by moving black shadows (which, in turn, are usually caused by flies). The frog's mental state could also represent *there is a black dot in the retina* and given that usually enough black dots in the retina are produced by flies, that could also explain why this device was selected for by natural selection. In a nutshell, the horizontal problem is that there are proximal and distal states of affairs and the correlation with many of them can explain the selection of the representational system (for a detailed mathematical model, see Martínez,2010).

42 Papineau (1993, p. 58-9, footnote 3) and Prinz (2000, p. 12) use the same names for these problems. These expressions are also used by Godfrey-Smith (1993), but my classification of the problem differs from his.

Concerning this problem, solutions among teleosemanticists differ. Neander (1995, 2006), for instance, thinks that the content the frog's representation should be somehow determined by the frog's discriminatory capacities. Given that frogs are only sensitive to black moving shadows, she suggests that the content of their mental state should be something like *there is a black moving shadow*. On the other hand, the indeterminacy problem has led Martinez (2010) to depart from teleosemantics and develop what he calls an 'Etiosemantic Account', which apparently solves the Indeterminacy problem by appealing to Homeostatic Property Clusters. The two views will be discussed in the next chapter (3.3.1 and 3.1.2). In my opinion, several reasons favor a different approach to this problem, more in accordance with a Millikanian version of Teleosemantics. My goal in this section is to show that there is a way First Teleosemantics can deal with this problem.

FIRST TELEOSEMANTICS AND INDETERMINACY There has been an intense debate in the literature as to whether teleosemantics is able to overcome these worries. Here I am going to argue that First Teleosemantics provides a solution to the horizontal problem. In fact, my answer does not differ from Millikan's original reply.[43] Nonetheless, I will show that Millikan failed to see that there is also a vertical problem of indeterminacy that needs to be addressed. I will then argue that First Teleosemantics has the resources for dealing with it as well.

First of all, we need to recast the indeterminacy problem in terms of First Teleosemantics. We need to identify a sender and a receiver which satisfy Sender-Receiver. In the frog case, the producer is identified with the visual system and the mechanism that consumes the representation is the tongue-snapping mechanism. On the other hand, the representation (the state that satisfies first Representation) is the frog's mental state that is produced by the visual system and elicits a certain response in the tongue-snapping mechanism.

How does First Teleosemantics deal with the horizontal problem? As we said, in the frog example the consumer system is the tongue-snapping mechanism. Now, consider the cases in which this system performs its function Normally; in these cases, the tongue-snapping mechanism yields a fly to the digestive system. Frogs do not digest black dots in the retina or black moving shadows, but flies. So, the particular condition that must figure in the most proximal and most comprehensive Normal explanation of how the *consumer* system performs its function is the presence of a fly (further specification will be provided below). Clearly, black dots in the retina or moving shadows are not what the consumer system requires in order to perform its function Normally (Elder, 1998, p. 352). According to this analysis, given that representational content is determined by the consumer system, we can exclude *there is a moving black shadow* as the content of the representation; the most proximal and comprehensive Normal explanation of how the tongue-snapping mechanism provides a fly to the digestive system mentions the presence of a fly and not the presence of a black thing, so the content of the representation is something like *there is a fly*. Compare it with the following fact: the most proximal and comprehensive Normal explanation of how the heart pumps blood

43 She has recently provided (what seems to be) a different reply to the indeterminacy problem (Millikan, 2004, p. 85). I think the previous reply is preferable.

does not mention the color of blood or where the electrical impulses come from.

Notice that there is no horizontal problem at the level of the function of the consumer. What determines the function of this consumer system is a purely causal mechanism. In the same way as the function of hearts is to pump blood, the function of the snapping mechanism is to bring flies (or perhaps *nutritious things*, see below) to the stomach. The horizontal problem arises when there is a relational function, that is, when the function is to do something when something else obtains. Then, the question of horizontal indeterminacy is the problem concerning the relata.

Let me rephrase the main idea of this section. The horizontal problem arises because there is a chain of states of affairs whose representation can account for the selection of the representational mechanism. Since moving black shadows usually correlate with flies (because flies usually *cause* the presence of a moving black shadow), systems that represent the former have the same fitness as systems that represent the latter (Martínez, 2010). A rough characterization of teleosemantics seems to leave this point undetermined: the representation of many states of affairs in the causal chain could account for the evolution of the representational system.

However, First Teleosemantics (and Millikanian Teleosemantics) is able to pick up one among the different causes of the representation: the content is the state of affairs that explains the success of the consumer system. Since, in the frog case, the consumer system is the tongue-snapping mechanism and this mechanism performs its functions Normally only if the frog catches a fly, the frog's mental state is representing (something like) *there is a fly.*

Of course, that would be the right conclusion if the only source of indeterminacy were the horizontal problem. But we will see in short that there is another objection lurking ahead: the vertical problem.

2.3.3.2 *The Vertical Problem*

Suppose we have got a theory that has a way of picking out the right state of affairs that should figure as the content of the representation. There still remains another indeterminacy problem, which consists in the fact that there are many properties that could be represented by the mental state. In other words, even if we restrict ourselves to Normal circumstances and to the state of affairs that caused the consumer system to perform its functions, there are many different properties that could do the trick. For instance, what the frog's consumer system needs are flies, but also a bunch of proteins, of food, nutritious stuff, of a fitness-enhancing thing, etc... and, of course, depending on the property that we pick up we will get a different content attribution: *there is a bug*, *there is food*,... how are we to choose among these contents?

Notice that, in contrast to the horizontal problem, the worry here does not concern different states of affairs located at various levels of distality in the causal chain (e.g. flies and black dots in the retina), but different properties that are causally relevant. So the fact that flies were nutritious, food, had certain proteins, etc... can explain why the consumer-system performed its other functions Normally.

Prima facie, the fact that the horizontal and the vertical problem have different origins suggest that we must look at different kinds of solutions. In contrast to what Millikan (1993) claims, I think it is

not obvious how First Teleosemantics can deal with this problem. Paying attention to the consumer system has helped us to pick out one among the different (distal or proximal) states of affairs, but it is not going to help us to zero in on one property. Certainly, the appeal to the consumption of the representation sows that black dots in the retina and black moving shadows can be excluded, but it is not going to pick out one among the different properties at the same level of distality.

The worry, then, is that consumer-based teleosemantics yields a too indeterminate content, because it is compatible with the content being *there is a fly, there is a bug, there are certain proteins,...*

2.3.3.3 *The Solution*

The previous discussion has shown that the version of the Indeterminacy problem that still needs a solution is the vertical problem. I argued that First Teleosemantics seems to be compatible with the content of the frog's mental state being *there is food, there is nutritious stuff, there is a fitness-enhancing thing,....* and it seems that teleosemantics is unable to determine which of these states of affairs is the content of the representation.

Fortunately, I think that this problem can also be solved by First Teleosemantics. In particular, the solution comes from focusing on the claim that the content of the representation is the state that must be mentioned in the least detailed and mostly comprehensive Normal explanation of how the consumer performed its function.[44] Let me elaborate on that point.

First of all, we saw that a Darwinian Population is a kind of Reproductively Established Family, i.e. it is composed by a set of individuals that tend to have certain properties in common in virtue of a causal process of copy. The Darwinian Population *fly*, in particular, consists of a set of bugs that tend to be nutritious, small, frog food and so on. Furthermore, there is a causal process (involving reproduction, stable environmental conditions, natural selection, etc..) that accounts for the fact that all flies tend to instantitate this set of properties. Consequently, *fly* is a Darwinian Population, in which their members tend to instantiate a set of properties (*being nutritious, being frog food, being small,...)* in virtue of some causal process of copy.

Now, if we accept that the kind *fly* constitutes a Darwinian Population, then there is a unifying reason why the properties that elicit the vertical problem were coinstantiated: the fact that the frog was confronted with a member of the Darwinian Population 'fly' explains that he got a small bug, a nutritious thing, a set of proteins, frog food and so on. In other words; among all the proximal explanations, there is one that accounts for the fact that *all* these diverse properties (being nutritious, frog food,...) were instantiated in Normal conditions: the fact that there was a member of the Darwinian Population 'fly'. Among all the least detailed explanation, appealing to the presence of a fly provided the most comprehensive explanation. Hence, the presence of a fly is the condition that should be mentioned in the relevant Normal explanation. Therefore, the content of the frog's mental states is *there is a fly around.*

Let me formulate my proposal in a different way. Remember that we said that, even if we focus on Normal conditions (those conditions

44 While my solution to the vertical problem differs from Martinez's (see 3.1.2) it has clearly been inspired by his proposal.

that were present in the circumstances that explain the selection of a mechanism), there are many different kinds of explanations. A *complete* Normal explanation mentions all facts that were present in Normal conditions. The *complete* explanation of how kidneys perform their function mentions the fact that gravity remained constant, the fact that the sun did not explode and also all the particular cells that have been involved in this process since the origin of kidneys. If we focus on the case of frogs, the *complete* Normal explanation of how the tongue-snapping mechanism achieved its function mentions the fact that there was a bug, a nutritious thing, a set of proteins and so on.

However, First Teleosemantics claims that we should pay attention to the least detailed and most comprehensive Normal explanation. So, among all the possible alternative explanations, the one that better satisfies the two desiderata of being the least detailed but most comprehensive explanation is the presence of a fly, that is, the presence of a member of the Darwinian Population *fly*. This is the fact that accounts for all the alternative proposals that appeal to some properties. There being a member of the Darwinian Population that we call 'fly' utterly explains the presence of a bug, of a nutritious thing, etc., so it accounts for the success of a tongue-snapping mechanism.

Therefore, the state produced by the snapping mechanism in frogs does not represent the presence of bugs, small things or nutritious organisms, but rather the presence of a fly. Consquently, the content of the frog's mental state is *there is a fly around.* This is, I think, an adequate content attribution, so it seems to solve the Adequacy Problem as well.

Let me point out that this reply to the indeterminacy problem does not entail that simple mental states always represent Darwinian Populations. In some cases the least detailed and most comprehensive explanation must mention the presence of an individual or a substance that grounds the coinstantiation of many properties, but which do not form a Reproductively Established Family (in the next chapter I will call them 'Instance-Types'. See 3.1.1). For instance, many organisms seem to be able to detect the presence of water. So one might raise the following (vertical) indeterminacy problem: are these organisms representing *there is water* or perhaps *there is a transparent substance*? or perhaps *there is a refreshing thing*? My claim is that there being water is the feature that must be mentioned in the least detailed and most comprehensive explanation, because it includes and accounts for the rest of candidates (*being transparent*, *being refreshing* and so on), and its being water is the key feature that accounts for all these features. After all, water is a transparent and refreshing substance. And notice that, in this case, water is not a Reproductively Established Family (different instances of water do not belong to the same kind in virtue of being copied from each other). This question will be discussed in more detail in 3.1.2.

2.4 Conclusion of Chapter 2

In this chapter I have offered a first version of a teleosemantic account, First Teleosemantics, which is intended as a semantic and metasemantic theory of content and representation. I carefully explained the notions involved and justified the conditions contained in First Teleosemantics. Furthermore, I have argued it is able to deal with the four main problems of previous theories. Obviously, a lot more has

to be done in order to show that this is the right account. Indeed, in the next chapter I will consider several arguments that show that First Teleosemantics needs to be amended in several ways. Nonetheless, I hope I have been able to persuade the reader that First Teleosemantics is a promising approach to the naturalization of representation and content.

3

IMPROVING TELEOSEMANTICS

The goal of this chapter is to develop, improve and defend the teleosemantic framework put forward in the previous chapters. I will discuss several objections to First Teleosemantics, some of which show that this approach needs to be amended in important respects.

The chapter is divided into two parts. In the first half, I will argue that there are certain ambiguities and limitations in some of the definitions provided in the previous chapter that indicate that the basic framework I oultined requires some mofidications. On the one hand, First Teleosemantics does not make the type/token distinction. On the other, First Teleosemantics cannot account for productivity, that is, for the capacity of certain mechanisms to produce new representations. I will spend some time introducing the notions that are required for an explanation of the type/token distinction and productivity and I will suggest slight modifications of some of the key definitions that were provided.

In the second part of the chapter, I will address the most important objections against teleosemantics (some of which I have already discussed in chapter 2). In considering these objections, I will compare my own account to Martinez, Neander and Shea's view, and I will argue that the (improved version of) First Teleosemantics is preferable. In a nutshell, this chapter is devoted to consider in detail several objections and ways in which First teleosemantics has to be improved.

3.1 TYPE/TOKEN

First of all, notice that First Teleosemantics fails to adequately draw a type/token distinction. That might have lead to some unclarities that need to be dispelled.

For example, First Sender-receiver was explicitly cashed out in terms of types. In turn, First Content appealed to systems and representations that *satisfy* First Sender-receiver, so First Content is likely to be naturally interpreted as being primarily applied to types. However, I take it that we want the teleosemantic recipe to be able to attribute semantic properties to particular representations (token), which belong to particular producer and consumer systems (token). So, we need to show how First Teleosemantics can be applied to tokens. Similarly, we should consider whether representations form REfs. We will see that an answer to these questions is not as straightforward as might seem at first glance.

Besides clarity, there is a theoretical reason for drawing a type/token distinction: as we will see later, distinguishing between types and tokens is going to be crucial in order to account for the productivity of certain representations and overcome some of the problems put forward in the literature. So let us consider how a type/token distinction can be introduced in our definitions.

3.1.1 Second Teleosemantics

In order to draw the type/token distinction, we need to explain what are types and in virtue of what property or relation do systems (token) and representations (token) belong to certain types. Thus, we need (1) a principled way of individuating types and (2) a recipe for classifying tokens into different types.

There are two interesting ways of individuating types that are relevant for us. I will call them 'REF-types' and 'Instance-types'.

REF-types First, we should recover the concept of *Reproductively Established Family* that was defined in 3.2.4. Reproductively Established Families (henceforth, REF), are composed of items that tend to resemble each other in several respects because they have been produced by some underlying causal process of copy. We defined them in the following way:

> **Reproductively Established Family** A group of individuals $d_1, d_2, d_3, \ldots d_n$ form a reproductively established family D iff $d_1, d_2, d_3, \ldots d_n$ tend to resemble each other in important ways because they are the result of some process of copy.

Usually, types are identified with REFs and tokens with the individual that belong to a REF. The inividual entities that belong to a REF are also called its 'members'. So, in some contexts, the distinction between types and tokens corresponds to a distinction between Reproductively Established Families and the individuals that form them. Therefore, on this first way of understanding the type-token distinction, a token d belongs to a REF-type D iff d is a member of the Reproductively Established Family D.

In 2.1.2 we saw that there is a set of REF that is specially interesting for our project: Darwinian Populations. Darwinian Populations will also play an important role in this chapter. Since they form REFs, the type/token distinction should be drawn in exactly the same way.

intance-types There is however a different way of classifying entities into types that is also important for our purposes. The color property of a rife tomato, the color property of blood and the color property of Red Lory birds (*Eos bornea*) all belong to the same type (redness) but they do not form a REF, since there might be no causal connection between them.

Of course, there is here a a whole range of interesting and complex metaphysical questions about the ontological status of this kind of types: are they universals or classes? Do particular color-properties belong to these types in virtue of instantiating these universals or in virtue of resembling each other to a certain degree? Can we define a satisfactory and objective notion of *instantiation* or *resemblance* that could play this role? These questions and many related issues are hotly disputed topics in metaphysics and I would like to remain neutral about them. For our purposes, it suffices if we say that these types are different from the types constituted by REF. Let us call them Instance-Types.

Having settled the distinction between two different sorts of types and relations between them and their tokens, we are now in a position to make a type/token distinction in a way that will allow us to attribute

content to representations (token). Let us present again the definition of sender-receiver structure given above (which it is not modified in any way, but I call 'Second Sender-Receiver' for simplicity) and modify First representation and First Content in order to make a type-token distinction available (lower case letters stand for tokens, uppercase letters for types):

Second Sender-Receiver

Any two systems P and C configure a sender-receiver structure if, and only if:

1. P and C have functions in accordance with Etiological Function
2. P and C have coevolved in such a way that a Normal condition for the proper performance of each is the presence and proper functioning of the other.
3. P has two functions:
   a) The non-relational function of helping C to perform its functions.
   b) The relational function of producing state R when another state of affairs S obtains.
4. The function of C is to produce an effect E. The least detailed and most comprehensive Normal explanation for C's performance of E involves S.

Second Representation

r is a representation iff

1. r is a member of the reproductively established family R.
2. R is a reproductively established family of states produced by a sender that satisfies Sender-Receiver.

Second Content

r represents s iff there are two systems p and c such that:

1. p and c are members of the Darwinian populations P and C, where P and C are systems that satisfy Sender-Receiver
2. r is a representation (in accordance to Representation) in virtue of being produced by p.
3. The least detailed and most comprehensive Normal explanation for c's performance of its functions when members of R obtain involves some members of (a REF-type or Instance-Type) S.

For simplicity, I will refer to Second Sender-Receiver, Second representation and Second Content as 'Second Teleosemantics'. Let me explain in some detail each claim contained in Second Representation and Second Content.

Second Representation introduces the type/token distinction. As we saw, a token belongs to a type either because this token belongs to a REF or because it is an exemplar of an Instance-type. Now, since systems that produce representations form REFs (they are reproductions

of each other), it is very plausible that representations also form a REF (they are reproductions from each other as well). If we focus on simple satates, it seems true that all representations tend to resemble each other in virtue of some causal process of copy. Thus, a particular state is a representation in virtue of belonging to a REF-type (see Millikan, 1984, p. 96-7). In other words, representations are states that tend to resemble past representations. I think the idea is essentially right, even though we will see that this condition will need to be slightly modified in order to account for more complex representations. Let us keep this simple formulation for the time being, until we get to the problems that will require an adjustment.

Let us consider now Second Content. The first condition just makes explicit an idea that was largely implicit so far, namely that particular representational systems have functions in virtue of belonging to certain reproductively established families (which, in this case, form Darwinian Populations). In the same way as particular hearts have a function in virtue of belonging to the Darwinian Population *heart*, representational systems have functions in virtue of belonging to a Darwinian Population.

Condition 2 merely applies Representation to a particular representation r contained within two senders p and c.

The most remarkable result of applying the notion of Reproductively Established Family to First Content in order to make a type/token distinction concerns condition 3 of Second content. Crucially, the state s (the represented state of affairs) can belong to a REF-Type or to an Instance-Type. For instance, beaver splashes can represent the presence of danger being around, but things that are dangerous need not belong to any reproductively established family (dangerous things need not be reproductions of each other). An organism might represent the presence of a red object, but red objects need not form a REF. So, while we require representations to form a REF, there is no restriction concerning the kind of states that can be represented. Prima facie, that seems to be the right result.

Nevertheless, the idea that that the represented state of affairs need not form REFs has been denied in Martinez, who has put forward a different teleosemantic account. I suggest to shortly discuss his own view on that matter.

3.1.2 *Martinez's Etiosemantics*

The idea that represented objects need not form REF contrasts with Martinez (2010)'s view. In order to solve the Indeterminacy problem Martinez claims that simple representational systems in cognitively unsophisticated organisms can only represent items that form Homeostatic Property Clusters (HPC). HPC are formed by a set of properties that are usually coinstantiated due to an underlying causal mechanism (Boyd, 1999a, 1999b). Paradigmatically, many natural kinds in biology form HPC. Natural kinds are composed of a set of properties that are coinstantiated because there is a causal mechanism that reliably produces this coinstantiation. For instance, horses are HPC because they possess a set of properties *(being four-legged, having a heart, being able to neigh,..)*, which are reproduced thanks to a set of underlying causal processes (stable environment, reproduction,...). Clearly, HPC qualify as reproductively established families, in our sense.

Martínez's view can be easily grasped by focusing on a particular example. Very roughly, Martínez's key insight is that representations R produced by the frog's hunting mechanism represent *there is a fly* because (1) historically, representations R (weakly) covaried with a set F of properties *(nutritiousness, flyhood, smallness,...)*, (2) there is an underlying causal mechanism that explains why this set F of properties was reproduced and increased the fitness of organisms that produced R, (3) the set F of properties and the underlying causal mechanism M correspond to the HPC that we intuitively call *fly.* Of course, this theory of content is accompanied with a metaphysical theory, according to which typically natural kinds (such as flies, horses or trees) are HPC.

First of all, let me point out that there is much I agree with Martínez. First of all, it is very plausible that many natural kinds in biology form HPC. Furthermore, this fact seems to partially explain the evolution of most, if not all, representational systems. The evolution of representational systems presupposes a certain stability in the environment, in the sense that certain regularity has to be in place in order for detection systems to evolve (Godfrey-Smith,1996). Otherwise, the pay-off of developing costly and (sometimes) misleading representational systems would be negative, and representational systems would have never appeared. This environmental stability can be ensured by the existence of HPC; instances of *fly* are instantiated around frogs because there is a causal mechanism supporting them (internal properties of flies, environmental features,...). Furthermore, the fact that this causal mechanism is in place partially explains the stability of the frog's environment, so that it contributes to the explanation of the evolution of the frog's representational system. Moreover, I showed that my solution to the Indeterminacy Problem was clearly inspired by this proposal.

However, adding HPC to the *conditions* for representing, as Martinez does, has some costs. Let me raise two objections against this proposal

INSTANCE-TYPES In particular, it has the striking consequence that only HPC can be represented by (relatively simple) representational systems. Some counterintuitive results follow from it. For instance, since the property *being red* does not form an HPC, but it is an Instance-Type (see above), then unsophisticated organisms cannot represent that something is red.

Let us consider in some detail some of these properties that intuitively do not form and HPC and see what Martinez can say about them. Scientists often claim that some organisms like toads or bees represent the presence of water (see Shettleworth, 2010, p. 516); so if Martinez wants to keep the truth of these scientific claims, he is committed to the claim that water forms an HPC. That is (to say the least) problematic, since there seems to be no underlying causal process that connects all instantiations of water, in such a way as to ground an HPC for water.

There are, of course, several replies available to Martinez. First of all, he could grant that there is no HPC that could be identified with water, but reply that in fact these organisms do not represent water, but *water in that pond*. Certainly, there is a causal mechanism that explains why water in this pond has certain properties, so it probably forms and HPC. The problem with this suggestion is that organisms seem to be able to represent things beyond *water in a pond;* bees, for instance, can represent the presence of water in causally disconnected places. For instance, if there is a source of nectar in an island in the middle

of a lake, they tend to avoid it. In this case, it does not seem that any individual pond is singled out, because it is very likely that the existence of the bee's representational system involves the presence of many different unrelated ponds and lakes. So it is very unlikely that the items that bees are referring to belong to an HPC of the form *water in that pond*.

Martínez (2010, p. 64-65), however, seems to favor a different solution to that problem:

> I the case of water it could 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 vapor. (...) So, the real kind closest to water that is an HPC whose specialized homeostatic mechanism is constituted by being H2O together with the water cycle. Maybe a not totally unfitting English name for such an HPC is *Earth water.*

But the claim that water forms and HPC is not devoid of problems. It seems that a consequence of Martinez's position is that if an organism perceives a certain amount of water that for some reason does not enter into the water cycle of Earth water, he is misrepresenting. For instance, perceiving some substance entirely composed of H_2O, which is contained in a deep cave that has never been in contact with the rest of water would constitute a misrepresentation. Similarly, perceiving some substance entirely composed of H_2O brought from another planet by a meteorite would produce misrepresentations. After all, this particular amount of water does not form and HPC with Earth water (or any subclass of Earth water, like *water in this pond*). These consequences are clearly counterintuitive.

More generally, the worry is that if HPCs are required for organisms to represent at all, it is hard to see how we can come to represent items that do not form an HPC. Either one is committed to the dubious ontological claim that all entities we can think of are HPCs, or one is committed to hold that many of what we think are contentful states, in fact, fail to refer. Neither of these claims is *prima facie* plausible.

PRODUCTIVITY There is a second set of problems. Martinez's theory not only requires that the represented state of affairs must be a member of a HPC; he also claims that this HPC must have coevolved with the representational system. In other words, he holds that the HPC that figures in the content of our representations must be part of the causal explanation of how the representational system evolved. However, if that were the case, it is not easy to see how we could come to represent new items, that is, items that were not present when our species evolved. More generally, the productivity of certain representational systems (the capacity to produce *new* contentful states) becomes problematic. Of course, I am not denying that an explanation within Martinez's framework is possible; perhaps the appeal to more sophisicated mechanisms or to Higher-order HPCs can do the trick (see Martinez, 2010). But things become much harder once we assume that the represented feature must be part of an HPC that has been causally relevant in the selection of the system.[1]

1 In behalf of Martinez's view, it must be said that, while his account makes it much harder to explain how organisms can represent things other than HPCs, it can straightforwardly

But, given that my own proposal is very close to Martinez's (specially in relation to the solution to the Indeterminacy Problem) one might wonder whether it does not fall prey to the same problems. Can Second teleosemantics (i.e.,Second Sender-Receiver, Second representation and Second Content) solve these two problems?

Concerning the first objection, there is a key difference between Martinez's and my solution to the Indeterminacy Problem: in contrast to him, my proposal does not require that the represented state of affairs forms a HPC. It is certainly true that in many cases what solves the problem is something like the existence of an HPC (more precisely, a Darwinian Population), but in other situations we might appeal instead to entities that do not form HPC.

Consider the two cases we have discussed so far. According to my proposal, the mental state of frogs is not indeterminate because the least detailed but most comprehensive Normal explanation of how the consumer performed its function claims that there is a member of the Darwinian Population *fly*. Nevertheless, in other cases there being an instance of *water* may provide the least detailed but must comprehensive Normal explanation (for instance, of why the consumer provided some refreshing, nutritious and potable substance to the stomach), even if water does not form and HPC. So my proposal can easily explain how cognitively unsophisticated organism can represent many entities that do not form HPC or REFs.

The second problem, however, is more puzzling. Is my theory well suited for explaining the productivity of some representational systems, that is, their capacity to produce new representations? It would be unfair to accuse other views of failing to account for the productivity of representational systems without showing whether my own approach can accommodate this fact. Indeed, I will argue that Second teleosemantics cannot explain productivity, but it can be modified in a way that can account for it. This is the issue I suggest to address in the next section.

3.2 productivity

I started this chapter by pointing out that First Teleosemantics does not make an adequate distinction between types and tokens. The second way in which First Teleosemantics (and Second teleosemantics) is unsatisfactory is that it is unable to account for the productivity of representational systems, that is, they cannot account for the emergence of *new* representations. This is a critical issue any theory of content must satisfactorily address, so I will spend most part of this chapter trying to show (1) why previous versions of teleosemantics fail to leave room for the existence of new contentful representations and (2) how this difficulty can be overcome.

3.2.1 *The Problem*

It is a platitude that some representational systems are able to create *new* representations. By 'new representation' I understand a represen-

account for the fact that most animals, including human beings, primarily represent HPCs.

tation that is endowed with a sort of content that is not shared by any representation produced by an ancestor of the organism.[2]

We can distinguish three different processes by means of which a system might be able to produce new representations:

1. There is a sense in which *indexical* representations generate new contents, and hence, generate new representations (Martinez, forthcoming). For instance, a frog's mental state represents the presence of a fly being *around now*. Thus, the same neurons going on at different times and locations will express different meanings. Similarly, we are familiar with plenty of expressions in natural language that express different contents depending on different parameters: utterer ('I'), time of utterance ('now'), day of utterance ('today'), and so on.

2. Organisms can also produce contentful *signs,* whit a shape (so to speak) that has never existed before. Suppose there is a particular color (e.g. Hume's infamous shade of blue) that has never been instantiated before and imagine George is the first to bump into an object that instantiates this color. George's perceptual state involves a sign that has never been tokened before and has an unprecedented representational content.[3]

3. Organisms can combine already existing representations in order to generate new (complex) representations with novel content. Human thought and language are paradigmatic examples. For example, it is probably the first time that anyone writes the following sentence: 'Thousand Fijian chicken sexers had dinner at Togo's best restaurant'. Nevertheless, this is a perfectly well-formed and contentful sentence that an competent English speaker can understand.

Cases of the third kind will be discussed in chapter 6, when presenting my own view on concepts. The task of this section is to account for 1 and 2, which are the first kinds of states that (phylogenetically and ontogenetically) produce new representational contents.

More precisely, I will first address the second kind of representations and then I will show how the account can be extended to the first sort of states. There are two main reasons for taking cases of type 2 as our starting point. First of all, indexical elements are usually unarticulated, i.e. there is usually no explicit part of the sign that corresponds with this indexical element that figures in the content. For instance, 'it is raining' means something like *it is raining here now*, but time and place are unarticulated in the sign. As a consequence, indexical elements are harder to identify. Secondly, any representation probably involves

2 New representations are the kind of representation that more starkly illustrate the problem, but the same kind of difficulties can be posed with certain kind of *old* representations. For instance, even if long time ago there was an organism that once had a representation with the same content as my actual mental state, First Teleosemantics cannot explain why the second time a representation is tokened it can be a contentful state. As we will see, the problem is that First Teleosemantics requires states and systems to be tokened many times before they can be said to be contentful. This is why new representations or representations that are tokened seldom are problematic.

3 On certain views, perceptual states work in the same way as indexical expressions, so according to them the example of the missing shade of blue should be classified as a particular case of indexical content (case 1). In 4.3.3 I will shortly address this question. In any event, here I am just trying to point at three different ways one can say that a representational system is productive; the particular examples I use are supposed to be merely illustrative.

some indexical features, so a progressive explanation (explaining how mechanisms X evolved from mechanisms lacking X) is much more difficult to carry out. Therefore, I will first explain how the capacity to produce new signs of type 2 could have evolved and how that process can be included in a proper formulation of teleosemantics. Afterwards, I will address the question of indexicality (type 1). Compositionality will be tackled in chapter 6 (see 6.1.1 and 6.5.1).

3.2.1.1 *Teleosemantics and Productivity*

As stated above, Second teleosemantics is unable to accommodate the existence of any of these three forms of new representations. The problems lie in condition 2 and 3 of Second Content and Second Representation.

Think first about condition 3. It asserts that the content of a representation is determined by a state that has Normally obtained in the past and must be mentioned in the least detailed and most comprehensive Normal explanation of how the consumer performs its functions. Now, if a state can only represent states of affairs that have been instantiated and played this role, then, by definition, representations can only represent states that have existed in the past. Only states that existed in the past can figure in evolutionary explanations. That prevents us from providing the semantics for 'now', 'today' or a representation of the missing shade of blue.

On the other hand, second Representation and condition 2 of second Sender-receiver claim that representations are states that belong to reproductively established families, which means that they share some properties in virtue of some process of copy. But it might not be obvious what reproductively established family new representations belong to, since by definition they are states that diverge in certain respects from past representations. Think about the perceptual representation of the shade of blue that has never existed in the past. Does this representation belong to the same reproductively established family as the representations of the rest of colors? In other words, is a perceptual representation of blue a copy of the perceptual representations of other colors (red, orange, yellow,...)? It seems it is not. But then, in which sense do they belong to the same REF? As I defined Reproductively Established Families, members of a REF must share certain properties. What are the properties that this representation of blue shares with past representations? Of course, we cannot appeal to the content of the representation, since the REF to which this state belongs needs to be established *before* we are in a position to attribute any semantic property. Hence, if we want representations to form Reproductively Established Families, REFs should be specified in more detail.

A more general worry is that, while it is quite clear how systems that produce a single representational state can evolve (Godfrey-Smith, 1996; Skyrms, 1996; 2010), it is not obvious how systems endowed with the capacity for generating new representations have emerged. How did *productive* representational systems evolve? Notice that if we get a plausible explanation of how such mechanisms could have naturally evolved, this approach might provide some illumination concerning the etiological functions of the systems, and in turn this explanation might help us to clarify the status of their representations and representational systems.

So these are the two main tasks of the following sections: (1) clarify certain conceptual issues concerning the capacity of producing new representations (2) finding a plausible hypothesis about the evolution of such systems and (3) amending Second representation and Second Content (specially conditions 2 and 3) in order to account for the productivity of representational systems. That settles the plan for the following section. First, I will present several tools that we need in order to formulate and address the question of productivity: open relational functions, mapping functions and consumption rules. Secondly, I will explain how representational systems bestowed with the capacity for producing new representations may arise. Finally, I will show how teleosemantics can accommodate these kinds of representations. Afterwards, I will present several ways in which mechanisms can evolve, which are able to combine representations and yield more sophisticated contents.

3.2.2 *Three Concepts*

In order to account for the possibility of new contents, we need first to introduce three different issues: the distinction between closed and open relational functions, the notion of mapping function and the concept of consumption rule.

3.2.2.1 *Closed and Open Relational Functions*

We saw in the last chapter that representational systems have what Millikan calls 'relational functions'. A trait has a *relational* function if the effect that explains why the trait was selected for is the production of something that bears certain relation to something else. Red-backed Salamanders have an internal mechanism that has the function of producing an internal sign *when* there is a predator around. Leopard frogs are supposed to produce an internal mental state *when* there is a fly around. In contrast, hearts have the function of pumping blood, period. That is the crucial difference between non-relational and relational functions.

We saw that having a relational function is a necessary condition for a system to be a representational system. We defined 'representation' as a state that stands between a sender and receiver system, and we saw that essentially senders have relational functions, so any representational system must have certain relational functions. But having a relational function is not a sufficient condition for representing. The chameleon's pigment rearranging device has a relational function, but it is not a representation because it lacks an adequate consumer system.[4]

In that respect, I think Millikan's (1984) notion of relational function was a great conceptual achievement in the naturalization of content. Nevertheless, I think she failed to introduce a crucial distinction between two kinds of relational functions: the functions had by systems

4 Similarly, Millikan claims 'Because it lacks an interpreter, the chameleon's color pattern, though it maps, is in no sense a 'sign'' (Millikan, 1984, p. 118) and 'A bee-flying in a certain direction is not an intentional icon, for there is no cooperating device that interprets it' (Millikan, 1984, p. 101). Nevertheless, let me point out that at some places of her (2004) she seems to have changed her mind and to be assuming that there is a sense in which signs can exist without an interpreter.

that can produce new representations and the function of systems that can not.[5]

Here is a case that is intended to highlight this distinction between two kinds of relational functions:

(A-bees and B-bees) A-bees and B-bees are two kinds of subspecies among bees. Both A-bees and B-bees use a similar device for communicating to other bees where nectar is. Once a honeybee finds nectar, it comes back to the hive and performs a waggle dance which indicates the position of resources to the other bees. However, evolution has produced a crucial difference between A-bees and B-bees: A-bees can only communicate two distances ('nectar at a short distance' and 'nectar at a great distance') and 4 directions ('direction of the sun', 'contrary to the direction of the sun', 'on the right hand-side of the sun' and 'on the left hand-side of the sun'). Obviously, these eight possible representations have all been tokened by past A-bees.

On the other hand, B-bees have a more elaborated system that enables them to represent much more precisely where the nectar is. The number of waggles, the direction of the dance and its intensity corresponds with a certain distance, direction and quality of nectar; but (and this is the curcial difference) their representations are not limited to a particular set. They can communicate, for instance, that the nectar is 247 meter away 56° on the right hand-side of the sun. Of course, with this sophisticated method, they can represent positions of nectar that none of the past bees have ever produced before.

I think the difference between the representational power of A-bees and B-bees is obvious. Let us call the relational function had by (a certain dance-producer-system in) A-bees '*closed* relational function', and the function of (a dance-producing-system in) B-bees '*open* relational function'. Now, for the reasons given above, Second Teleosemantics can only explain the representational capacities of A-bees. We need to leave room for systems with open relational functions.

In order to carry out this discussion, it might be useful to introduce two other technical notions put forward by Millikan (1984) and Godfrey-Smith (1996).

3.2.2.2 *Mapping Functions*

Mapping functions are functions in the mathematical sense (in order not to be misled by functions in the sense of Etiological function, I will usually refer to mapping functions as 'M-functions', where 'M' stands for 'mapping' and 'mathematical function'). More concretely, a mapping function is a (mathemathical) function that in Normal circumstances holds between representation and representatum (Millikan, 1984). So, the domain of a mapping function is a set of representations (i.e. a set of vehicles of representation) and its co-domain is a set of objects, properties, relations and states of affairs. The key idea is that

5 Of course, it might well be that *in fact* all representations have an indexical element, and hence that all representations are of the first kind. But still, the two kinds of functions are conceptually different and require different settings in order to evolve (see below).

when a given representational system evolves in the way stated by Second Teleosemantics, a certain mapping function between representations on the one hand and represented objects, properties and states on the others, is established.

For instance, there is an M-function that Normally maps the chameleon skin onto the surface color of the thing the chameleon is sitting on. In that case, the M-function could be defined as follows: f(x) = x, which associates the color of the chameleon skin with the color of the surface.[6]

Notice that M-functions can be as bizarre as you like; for instance, the open relational function of the chameleon's pigment-rearranging device could have been 'produce a color that is slightly darker than the color of the surface you are sitting on', or 'produce the color that is opposed in the spectrum to the color of the surface you are sitting on'. This is important because different representational systems specify different M-functions between representations and representata. At the very end, this is what will allow us to account for the extreme variety of contents that can be determined by different representational systems.

3.2.2.3 *Consumption Rules*

There is a second kind of rules that are relevant in this context, which Godfrey-Smith (1996, p. 181) calls 'consumption rules'. As we said, a mapping function is a mathematical function that (at least in the cases considered) maps representations onto states of affairs that the consumer needs in order to perform its functions in a Normal way, as stated in second Sender-Receiver and Second Content. Consumption rules are also mathematical functions but, in contrast to mapping functions, they are determined by certain dispositional properties: consumption rules determine a function that maps representations (i.e. vehicles of representation) onto particular behaviors that are elicited on the consumer system (assuming the consumer is not broken or damaged). That is, there is a function such that, given a certain representation as argument (e.g. bee dance nr. 857) it yields a certain behavior of the consumer system as value. This is the behavior that bees are disposed to perform given this representation. Consumptions rules are determined by what the consumer has the disposition to do if not broken or damaged, rather than what the representation is supposed to map onto. For example, the consumption rule of the snapping mechanism in the frog prey-systems is extremely simple; it maps activation in a certain set of neurons onto the throwing of the tongue.

Having defined the notion of open and closed relational functions, mapping function and consumption rules, let me provide a plausible story about how closed and open relational systems arise. Remember that sketching plausible and evolutionary processes is important because, according to First Sender-Receiver and second content, whether a state qualifies as a representation and what its content is depends on the evolution of this kind of representations. So in order to explain the distinction between mechanisms that can generate different sorts of contents it will be very useful to have a plausible story about the evolution of these mechanisms and show why they are significantly different. Hence, let me address a particular case of evolution that

6 While there is a mapping function between the chameleon skin color and the color of the surface it is sitting on, remember that the chameleon's pigment-rearranging device is not a representational system, because it lacks a suitable consumer-system.

might illustrate the distinction between closed and open relation functions, mapping functions and consumption rules. We will draw several conclusions from it.

3.2.3 *The evolution of systems with open relational functions*

Here is an example of the kind of process that might generate open relational functions (of course, other processes are also possible).

Suppose an organism evolves a representational mechanism like any of the frog's or beaver's warning signals we consider earlier. In all these cases, the producer system P can at most produce a limited set of representations (say *r* and *r'),* which are supposed to correspond to very limited state (say, *s* and *s'* respectively) in accordance with Second Teleosemantics. The beaver's tail splashing the water can only mean *there is danger around.* There is no way it can also indicate the kind of danger or its direction, for instance. In this situation, the sender produces representations that are supposed to correspond to certain states in accordance with a mapping function *f.* In particular, the domain of *f* is r and r' (let us call this set 'R') and the co-domain is S (for s and s'). The mapping function *f* can be defined in the following way: if its argument is r, its value is s and if its argument is r', it yields s'. This is a simple and clear example of closed relational function.

Now suppose that one day, a certain organism endowed with this representational system produces a new representation *r''*[7] and suppose that the fact that a state *s''* obtains causally explains why the consumer system performed its functions. Indeed, we can imagine that this particular organism produces many different representations r'', r''', r'''',.. and usually enough the ensuing behavior (b'', b''', b'''',..) increases the chances of surviving in response to a different environment (s'', s''', s'''',...). Now, if after many generations in which this set of states leads to successful behavior we ask what explains that this mechanism has been selected for by natural selection, the most proximal and comprehensive Normal explanation appeals to a rule with certain variables; in the same way that what explains the pigment-rearranging device in chameleons is not that it produced green color, or blue color, but the production of the color of the thing the chameleon was sitting on, the most proximal and comprehensive Normal explanation of how this representational mechanism was selected for must mention the fact that the producer system produced a set of states R (r, r', r'', r'''...) that were supposed to correspond with a certain state of affairs in accordance with a mapping function. So to speak, it is a mapping function that is selected, rather than a particular state. This is how open relational functions might arise.

But, one might worry, how can it happen that, by looking at *r''*, an organism reacts successfully to *s''*, given that *r''* has never been performed in the past? Surely (the objection runs) the conditions in which r'' was tokened were entirely abNormal. How can an organism react in an appropriate way to a certain abNormal condition? This is a very reasonable doubt concerning the explanation of the evolution of open relational functions.

The most plausible way open relational functions can evolve is by riding piggyback on an existent system with a (closed) relational function. Again, let us go back to the beginning of the story; suppose that a

7 We will clarify below in what sense a new state can be said to be a representation.

mechanism with a closed relational function determines an mapping function f that only maps r onto s and r′ onto s′. Then, one might suppose, even if r″ has never been produced before, r″ can *cause* the consumer system to adequately react to s″, given the consumption rules that are in place. Given the way the system is build (i.e. given certain consumption rules), producing r″ causes the organism to reacting a certain way, which happens to be fitness-enhancing. In other words; the production of r″ (which, as I said, has not been produced in the past) might cause the consumer system to react adequately to s″ because this is what the consumer system is disposed to do given r″. The consumption rules that are in place help to increase the probability of a successful reaction by the consumer system, even if the situation is abNormal.

Let me illustrate this point with the following case:

(Kimus) Kimus are small animals that live inside caves. They are preyed by snorfs, a large red-skinned carnivore. Since kimus live inside caves, they can only see the snorf's red skin either when snorfs graze at the entrance of the cave (where there is plenty of light) or when snorfs are very close to kimus inside the cave (there is very little light inside the cave). In particular, when snorfs are somewhere in the middle of the cave, kimus fail to see them. Further, suppose that, during million of years, kimus have evolved a very simple representational mechanism, which consists of only two representations, r and r′. When a kimu perceives a small red figure (Normally, when snorfs are at a long distance, roughly at the entrance of the cave) it produces an internal sign r, which causes him to slowly move deep into the cave (kimus are slow creatures and need a lot of energy in order to move fast. So, unless it is completely necessary, they prefer to move slowly). In contrast, when they perceive a big red figure (Normally, when snorfs are in the vicinity so that, despite the little light, kimus can see a large red object), an internal sign r′ causes them to quickly move deep into the cave. Accordingly, it seems 'r' means something like *there is a snorf far away* and r′ *there is a snorf at a short distance.*

Now, one day a family of kimus K moves into a cave that is very well illuminated. As always, when kimus perceive a small red figure, r is activated, which causes them to slowly move into the cave, and when they perceive a big red figure they move quickly. However, due the the unusual light conditions of the cave where family K moved in, snorfs at a medium distance also produce a certain representation in kimus. When a snorf is inside the cave (but not too close), it produces an internal state r″ (and activation that lies between r and r′), which has never been produced in the past. Since r″ produces a neuronal activation between r and r′ (it is caused by a medium-sized object), when r″ is produced, kimus happen to move at a half-speed into the cave.

Producing r led to slow movement; r′ led to quick movement; so one can understand why r″ (a medium activation) caused the consumer system to react by causing the kimus to move into the cave at a half-speed. That

is what followed from the consumption rules of the kimu's consumer system.[8] Nonetheless, notice that the first time a kimu produces r″, it is an abNormal situation. At this point r″ was a misrepresentation, since the receiver was not designed to react in any way with respect to r″; this sort of representations had never existed in the past. Despite this fact, given the way the consumer system was supposed to react to r and r′ (so, given c's consumption rules) r″ prompted successful behavior.

Now we are only a small step away from the emergence of an open relational function. Suppose that in this enlightened cave different neuronal activations are produced; each time a snorf is at a different distance inside the cave, different states are generated: r″, r‴, r⁗,.. and, of course, usually enough the ensuing behavior leads to an increase in the chances of surviving. Now, if after many generations in which this set of states leads to successful behavior we ask what explains that this mechanism has been selected for by natural selection, the most proximal and comprehensive Normal explanation of how the kimu's representational mechanism was selected for must mention the fact that the producer system produced a set of states R (r, r′, r″, r‴...) that were supposed to correspond with a certain state of affairs (s, s′, s″, s‴,.. which correspond to snorfs being at different distances) *in accordance with a complex mapping rule.* In that case, the primary explanation does not need to mention the particular states that were produced, but the mapping function that was selected. Which particular states were produced is largely irrelevant (in the same sense that which particular colors have chameleons produced in the past is largely irrelevant in order to explain the functioning of its pigment-rearranging mechanism). The key element is that a certain mapping function is selected, which maps states r* onto states s*.

More schematically, we can depict the historical situation that gives rise to open relational functions in the following way:

- A certain representational system with (closed) relational functions evolves. It works according to a certain (simple) mapping function *f*, which maps r onto s and r′ onto s′.
- A sender p produces a new state r″ which, given the previous consumption rules, causes the consumer system c to adequately react to s″. The fact s″ obtained and c performs its functions (this time, though, in an abNormal way).
- Multiple representations are produced that did not existed in the past, but that prompt successful behavior given the existent consumption rules. Several generations pass by.
- At some point, the most proximal Normal explanation of how the consumer system performed its functions mentions the fact that a set of representations R mapped onto a set of representata S in accordance with a M-function that can generate new representations.

8 Of course, this story requires that the way representations r and r′ are implemented admits of degrees between them. That is, if instead of being weak and strong neuronal activations, r and r′ correspond to neuronal activations at different and unrelated places of the brain, there might be no way to find a middle activation r″ that elicits the right behavior. In that case, the consumption rules (the mathematical function that maps representations onto behaviors) will be different.
Notice, however, that I am here merely describing a possible way productive representational systems might evolve.

As I said above, there might be other processes by means of which systems with open relational functions might evolve. My goal was only to explain how representational systems capable of producing new representations might have evolved.

3.2.4 *Productivity and* Second Teleosemantics

The last issue I would like to turn our attention to before reformulating Second Teleosemantics is the problem with Second representation and condition 2 in Second Sender-receiver. The question here is how can new states be considered representations if they have not been copied from past representations.

Second representation claims that, by definition, representations are states that form REFs, that is, sets of entities that tend to have certain properties in common because they are the result of some causal process of copy. However, since new representations have never been tokened before, it is not clear to what extent new representations are a copy of past ones. Is bee dance nr. 463 a copy of bee dance nr. 14? Is the chameleon's unprecedented skin color blue a copy of the past skin colors red, green and so on? I think the best way to accommodate this sort of cases if by making a distinction between two kinds of reproductively established families. More concretely, we have to specify two different ways in which a process of copy can give rise to a significant set of reproduced entities.

Indeed, in the first definition I provided I said that a REF is composed of a set of states that are copied from each other, but the notion of *being copied from* was left largely unexplained. If we spell out this processes of copy in two different ways, we might be able to draw the relevant distinction between two different kinds of REFs.

Millikan (1984, ch.1) already distinguished two different kinds of reproductively established families (even if she used this distinction for a different purpose and formulated it in a slightly different way). Following her I will call them first-order and second-order reproductively established families:

First-order Ref A set of individuals $d_{1,}, d_{2,}, d_3, ..., d_n$ form a reproductively established family D iff

1. There is a set of properties $F_{1,}, F_{2,}, F_3$ such that $d_{1,}, d_{2,}, d_3, ..., d_n$ tend to instantiate a high number of these properties
2. For any d, the fact that d's ancestors had $F_{1,}, F_{2,}, F_3, ..$ in part causally explains why d has $F_{1,}, F_{2,}, F_3, ..$

Genes, for example, belong to first-order reproductively established families, because they tend to have some properties in common due to fact that each gene is always produced by a previous gene. This is the sense of 'reproductively established family' we have mainly been working with. But there is a second way things form families that also deserves careful examination:

Higher-order Ref A set of individuals $d_{1,}d_{2,}d_{3,...}d_n$ form a higher-order reproductively established family D iff it is a function of a device that belongs to a *first-order* reproductively established family to produce them.

For instance, there is an important sense in which a set of toys produced in a factory form a family - they are members of a higher-order reproductively established family, because they are produced by a device whose function is to produce them.[9] More generally, any set of items that are produced by a mechanism whose function is to produce them forms a higher-order reproductively established family.

Now, when we turn to certain biological items like *traits*, things become a bit more complicated. Certainly, traits such as lungs or the eyes form higher-order reproductively established families, since there is an item whose function is to produce them: genes. So, surely, those traits form higher-order REFs. But notice that they also form first-order REFs (Godfrey-Smith, 2009, 67-89). First of all, all lungs tend to resemble each other. Secondly, they tend to resemble each other because lungs have certain positive effects on organisms, what causally explains why lungs tend to have the same properties. Put in a different way: traits form Darwinian populations, and given that Darwinian populations are first-order REFs, traits form first-order REFs.[10]

This issue is going to be important later 3.2.6, since I will argue that it is only in virtue of belonging to a first-order REF that items can acquire functions. As we will see below, that excludes what Millikan calls 'derived proper functions', which according to her are functions that some items acquire in virtue of being traits of a higher-order REF.

It is time focus on Second Teleosemantics and modify the theory in a way that includes all the points made in this chapter.

3.2.5 *Reformulating the theory*

I previously argued that in order to account for the productivity of representations (mainly in the sense of 1 and 2 above) we had to

9 In contrast to Millikan, I think (and will argue in 3.2.6) that only first-order reproductively established families can ground function attributions. If this is right, her claim that devices that *have the function* of producing members of higher-order reproductively established families form a first-order reproductively established family is redundant; since they have the *function* to produce ds and only devices that belong to first-order reproductively established families can have functions, they must belong to first-order reproductively established families.

10 Since Millikan formulates first-order REF in a slightly different way and in order to fulfill a different task, she thinks traits do not form first-order REFs:

> Biological devices such as dogs, human hearts, and stickleback fish's mating dances are not reproductions of one another either. (...) It is not directly because his father danced so that the stickleback dances so. Had his father been injured and hence danced differently, that would not have caused him to dance differently. It is not directly because my father and mother had two legs that I have two legs. Mutilated parents can produce normal children. Wooden legs are not inherited. Rather, the stickleback's *genes* and my *genes* (tokens) were reproductions of earlier gene tokens harbored by our respective parents, and similar genes produce similar parents. (Millikan, 1984, p. 22)

Certainly, this is the result of her definition of first-order and second-order REF, which slightly diverges from mine.
I have relaxed the notion of first-order REF for two reasons. First, I think there is an important sense in which the properties of a trait are causally explained by the properties of its ancestors, while there is not a sense in which the properties of a toy produced in a factory are causally explained by the properties of previous toys produced in the same way (of course, I am supposing that in this factory new toys are not modified in light of the properties of previous toys). This is one of the reasons lungs are selected for and toys are not (Godfrey-Smith, 2009, 67-78). In contrast, on Millikan's notation, traits and toys are classified under the same label. Secondly, Millikan thinks both first-order and higher-order REFs can ground attributions of functions. I will argue that only first-order REFs can, so it is crucial that it contains all and only those members that acquire functions in the relevant sense.

introduce systems with 'open relational functions' and explain how they could evolve. Furthermore, we saw that accounting for the existence of open relational functions requires certain adjustments in Second Teleosemantics. It is time to address the previous formulation of the theory.

First, there are two aspects in which second Sender-Receiver has to be improved. First of all, we need to specify that representations can belong to types in virtue of belonging to a higher-order REF. Of course, they can also form a first-order REF (this is the case of the frog's representations, for instance) but in many cases, belonging to a higher-order REF suffices.

Secondly, we need to introduce the notion of mapping function, which was implicit in the previous definition. We need to make it explicit due to its crucial importance in order to account for productivity (see Shea, forthcoming).

Consequently, we get the following result (italics mark the changes from past versions):

Third Sender-Receiver

> Any two systems P and C configure a sender-receiver structure if, and only if:
>
> 1. P and C have functions in accordance with Etiological Function
> 2. P and C have coevolved in such a way that a Normal condition for the proper performance of each system is the presence and proper functioning of the other.
> 3. P has two functions:
>   a) The non-relational function of helping C to perform its functions.
>   b) The relational function to produce a *set of states R*, which are supposed to map onto another *set of states S in accordance with a certain mapping function* f.
> 4. The function of C is to produce an effect *(or set of effects)* E. The least detailed amd most comprehensive Normal explanation for C's performance of E involves *members* of S.

And we also have to modify Second Representation and Second content:

Third Representation

> r is a representation iff
>
> 1. r is a member of the *higher-order* reproductively established family R.
> 2. R is a reproductively established family *in virtue of being produced* by a sender that satisfies Sender-Receiver.

Third Content

> r represents s iff there are two systems p and c such that:

1. p and c are members of a Darwinian Population[11] P and C, where P and C are systems that satisfy First Sender-Receiver and Darwinian Population.
2. r is a representation (in accordance with Third Representation), in virtue of being produced by p. [12]
3. s is the state that r is supposed to map onto in accordance with *f*.

There are two important aspects of this definition. First, traits (systems p and c) belong to Darwinian populations, that is, first-order REFs, while representations (r and s) belong to higher-order REFs. That captures the idea that traits are selected for, and hence (in accordance with Selection for) they are copied from each other. In contrast, representations are not selected for; if they were, new representations would be excluded from Third Teleosemantics and thus it would not be able to account for productivity. Representations form a (higher-order) REF in virtue of being produced by a mechanism whose function is to produce them (in the same way toys form a higher-order REF because they are produced by the same model). Of course, they might also form first-order REFs, but this is not needed. Thus, by exclusively requiring that representations form a higher-order REF, we can include all representations produced by senders (e.g. a new bee dance, a new skin color) within the definition.

Secondly, notice that s (the 'referent' in Millikan's (1984, p. 113) terms) need not exist or have existed in the past. This is also required by productivity: a new representation can map onto something that has never existed or will never exist. What needs to have existed is an M-function *f* such that, given a certain representation r, it yields a single referent s that r is supposed to map onto.

This is how the two challenges that productivity raised against Second Teleosemantics can be overcome by Third Teleosemantics: the question of how *new* states can be represented if content is determined by past instances, the issue concerning the representational status of new representations and the the question of how all these systems can evolve. I think all three questions have been solved in the discussion leading to Third Teleosemantics.

3.2.5.1 *Indexicals*

Now, what about indexical elements? We saw that there is a sense in which any system that produces indexical representations is also productive. Can this indexical productivity be explained by Third Teleosemantics? I think so. As I said earlier, mapping functions can be as bizarre as one likes. So in order to account for the fact that some mechanisms are able to produce states that change their meaning depending on a certain parameter (time, location, utterer,...) we just need to assume that there is a mechanism that produces representations that are supposed to map onto the world according to this kind of mapping function. For example, the mapping function that governs the beaver's alarm mechanism maps beaver splashes onto the presence of danger

11 That is, p and c are first-order reproductively established families.

12 Since in many cases higher-order reproductively established families also form first-order reproductively established families, condition 2 includes cases where a mechanism's function is always to produce the same simple state r (e.g. the frog's mental state when catching flies).

now and *here*, because in Normal conditions, what explains the success of the consuming mechanism is the presence of danger *at the same time* and *at the same location* of the beaver splash. There is nothing mysterious about indexicals; they can be perfectly accommodated within Third Teleosemantics.

3.2.5.2 *Kripkenstenian Worries*

I have assumed that there is a mapping function between the representations and the represented states that accounts for the productivity of some representational systems. Now, someone might raise the sort of troubles that Kripke (1982) formulated in this famous book on rule-following: for any two finite sets of entities, there is an infinite set of mathematical functions between them. So, if we take the set of all past states produced by senders of some kind (i.e. the set of *representations* of a certain kind) and the set of all past states that explain the success of the consumers of a certain kind (i.e. the set of *representata*), there are infinite mapping functions between them that could explain the selection of the mechanism. So, unless we specify a criterion for picking up a particular mapping function, it seems that there is a huge indeterminacy among mapping functions. The theory seems to be impotent in order to select one mapping function rather than another. For instance, what determines the fact that the pigment-rearranging device is following the rule *produce the same color of the thing the chameleon is sitting on* and not the rule *produce the same color of the thing the chameleon is sitting on if it is not blue*$_{341}$*; in that case, produce black skin?*

A full response to this issue would require a long discussion that we cannot develop here. However, notice that the most natural reply is that we should pick up the mapping function that must be mentioned in the least detailed and most comprehensive Normal explanation; and the least detailed Normal explanation will appeal to the mapping function that does not mention exceptions of the kind *'and if it is blue*$_{341}$ *produce black skin'*. The least detailed and most comprehensive Normal explanation is the one that is able to explain a preponderant number of past cases mentioning the less number and more comprehensive features (2.3.3.2). Furthermore, remember that appealing to this explanation is not arbitrary. There is supposed to be a causal process that grounds this explanation. So this solution to Kripkenstenian worries would not be arbitrary either.

Summing up, I think Third Teleosemantics is about the best we have for naturalizing perceptual and conceptual content. I think we have set up all the conceptual tools we are entitled to use in order to provide a naturalization of intentional content. Given the magnitude of the project, Third Teleosemantics might look too simple an account for dealing with complex mental states such as concepts. The goal of the rest of the dissertation (specially part II) is to show that these doubts are unfounded.

Indeed, most teleosemanticists think Third Teleosemantics (or something along these lines) is insufficient for the naturalization of intentional content. In particular, Millikan and others following her (Shea, Papineau,...) maintain that there is another tool that is legitimate and required for dealing with complex representations: what they call 'derived functions' and 'adapted functions'. Hence, before moving to the objections to Third Teleosemantics, let me argue why I think these

notions cannot contribute in any relevant way to the naturalization of intentional content.

3.2.6 *Adapted and Derived Functions*

In her naturalistic account on content, Millikan has extensively used two additional notions: adapted and derived functions. These have usually been accepted by people following her (Elder, 1998; Sinclair, forthcoming; Preston,1998; Papineau,2003). I would like to examine these two concepts, and see whether they provide any substantive contribution to the toolkit elaborated so far. I can advance that my conclusion will be negative.

3.2.6.1 *Adapted Functions*

The notion of adapted function derives from the previous notion of relational function, defined in 2.2.3. As we saw, a system has a relational function if its function is to produce something that bears a certain relation to something else. The paradigmatic example is the chameleon, whose pigment-rearranging device has the relational function of producing a skin color that matches the color of the surface it is sitting on. Now, in a given occasion the function of the mechanism is going to be satisfied by the instantiation of some particular color (green, say). Millikan calls the state of affairs the mechanism is reacting to the 'adaptor' and claims that once the mechanism has an adaptor, it acquires an *adapted function*.[13] If the chameleon sits on a brown branch, the adaptor is the particular brownness of the branch and the chameleon pigment-rearranging device acquires what we can call the *adapted relational function* of matching the brown surface. In Millikan's terms:

> When a device has a relational proper function, given some specific thing that the device is supposed to produce in relation to, the *device* acquires what I will call an *adapted relational proper function.* (...) Once a bee has spotted nectar at a particular place, the dance choreographic devices in the bee acquire as an adapted proper function the production of a specific dance. (Millikan, 1984, p.40. Emphasis added).
>
> An adapted proper function is a relational proper function adapted to a given context. (Millikan, 1984, p.42)

Crucially, notice that the device with a relational function and the device with an adapted function are the very same device. Furthermore, there has not taken place any process of selection that could endow this mechanism with a new function. Hence, a reasonable hypothesis is that 'Adapted relational function' is just a shorthand for 'relational proper function *plus* adaptor' (see Millikan, 2002, p. 125). Strictly speaking, then, adapted relational functions are not new functions of the mechanism, so they do not introduce any new feature into the picture. They should be regarded as a mere notational variant.

13 Again, Millikan's own term is 'adapted proper function'. For simplicity and in accordance with the perspective given in the first section of chapter 2 (see **??**), I will refer to Millikan's proper function as 'functions'.

3.2.6.2 *Derived functions*

The notion of derived function is less clear than that of an adapted relational function. Millikan defines this notion in the following way:

> 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 the *products* of devices with relational functions (i.e. a particular brown skin produced by the chameleon, or bee dance nr. 879) also have some kind of function: derived functions. A clear example is the derived function of a chameleon's brown skin, which (according to common wisdom) is to camouflage the chameleon.

Notice that, in contrast to adapted functions, which are had by the very same mechanism that has the relational function, derived functions are possessed by produced items, which might not have been selected for. Derived functions are possessed by new items (particular skin colors, particular dances) in virtue of being the products of functional devices (see below). Hence, while the introduction of adapted relational functions was a mere terminological move, derived functions do not seem to be reducible to relational functions. As a result, if these functions exist, they might constitute a new and fruitful notion for naturalizing conceptual content.

Millikan distinguishes two kinds of derived functions: relational and non-relational derived functions. Furthermore, since an adapted function is just a relational function plus an adaptor, devices with derived relational functions can also have derived adapted relational functions.[14]

An example might be useful here; if a bee finds nectar at spot L, and has a mechanism M which produces a dance D so as to bring other bees to the nectar, then we can distinguish the following functions:

1. M's *non-relational* function of producing something that allows the consumer system to perform its own functions (see condition 3a in THIRD TELEOSEMANTICS , in 3.2.5).

2. M's *relational* function of producing particular dances that are supposed to map onto locations of nectar according to a certain mapping rule f (see condition 3b in THIRD TELEOSEMANTICS, in 2.1.2).

3. M's *adapted relational* function of producing D, whose adaptor is nectar at L.

4. D's *derived non-relational* function of bringing other bees to gathering nectar.

5. D's *derived relational* function of bringing other bees to a particular source of nectar.

6. D's *derived adapted relational* function of bringing other bees to nectar at L.

14 Millikan (1984) vacillates between calling them *derived adapted* proper functions (p.43) or *adapted derived* proper functions (p.42)

We saw that adapted functions do not introduce any substantive entity; it is just a different way of saying that a mechanism has a function and a certain adaptor. Accordingly, 3 does not provide any new function; 3 is just the claim that 2 holds and there is certain adaptor (nectar at L). Similarly, 6 is not attributing any new kind of function; 6 merely states that 5 holds and there is an adaptor. So only 4 and 5 seem to be ascribing new kinds of functions.

Therefore, our next question is whether we can make sense of the notions of derived non-relational and derived relational functions. I we get an affirmative answer, we should assess their role within a theory of the content of representations.

3.2.6.3 *Are there derived functions?*

A first worry we might have with the notion of derived function is that it seems to be in conflict with the general approach to intentionality suggested by teleosemantics. The idea that some devices have functions merely in virtue of being produced by other traits seems to contradict the key insight of Etiological function, since a necessary condition for a device to have an etiological function is that it be selected for. Why should we think particular bee dances can acquire functions just in virtue of being produced by a dance-producing mechanism? How could there be such a *transfer of normativity* from the producer to the product (Martinez, 2010, p.83)?

I think it is not clear how Millikan would answer that question. On the one hand, she sometimes claims that derived functions are passed on from the producers to their products, even granting that the products are not selected for.[15] In other places, she argues that there is nothing in the account of derived functions that contradicts her previous analysis.[16] Be as it may, I think it should be clear that the idea of a function being passed on is mysterious and threatens to undermine Etiological function. As Preston, (1998, p. 234) suggests:

> To put it bluntly, the introduction of derived and expanded proper functions ['expanded functions 'are a kind of derived functions] means that proper function in general does and does not essentially involve a selection history in the primary biological sense, and consequently it both is and is not normative.

Now, in chapter 2 (see 2.1.2) I painstakingly argued that the etiological theory of function is the only one that can account for an attribution of functions with normative import. And since Etiological function requires functional states to be selected for, I think it has been already

15 For instance, consider the following quotes: "This is the doctrine of derived proper functions, in accordance with which certain kinds of teleofunctions that are built into an animal during evolutionary history interact with the environment of the animal to produce *new teleofunctions, new biological purposes for those individuals, without the mediation of additional selection processes*" (Millikan, 1993, p. 133, emphasis added) "Thus, it happens that artifacts have as derived functions the functions intended for them by their makers"(Millikan, 1998, p. 204)

16 Consider, for instance, the following quote: "It was to simplify the description of these complex relational structures and processes that I introduced in LTOBC the terminology 'adapter proper function' and 'derived proper function'. I intended this merely as a useful nomenclature. It is not an addition or set of extra clauses widening or narrowing the original definition of 'proper function', but merely a way of talking more easily about phenomena that had already been captured by that notion, given that traits and mechanisms can have relational functions. (Millikan, 1998, p. 201)

shown that particular bee dances or particular skin colors cannot acquire a function merely in virtue of being the product of a trait with certain functions. Functions are not cheap.

So, should we just dismiss talk of derived functions and reject it as a confused notion? I think that would be too quick. In fact, I think there is a way of making sense of the idea of new and particular representations having functions, which is compatible with the teleosemantic framework. The idea parallels the kind of analysis I provided concerning the function of hearts; even if hearts are produced by genes, whose function is to produce them, we do not need to assume that hearts have functions *in virtue of* being the products of genes, because hearts themselves are selected for. Similarly, even if bee dances are the products of certain functional mechanisms, we do not have to accept that particular bee dances have functions *in virtue of* being the product of dance-producing mechanisms; we can instead defend that bee dances have functions in virtue of being selected for as such. In fact, in some places Millikan herself seems to accept this kind of explanation, which makes derived functions compatible with ETIOLOGICAL FUNCTION:

> Is it also a proper function of the dance itself to produce this direction of flight? The answer may at first seem to be "no", for it seems theoretically possible, at least, that the particular bee dance has no ancestors. (...) Then the 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.
>
> But this overlooks a principle that is fundamental. (...) What is of interest is whether there is *a* sameness among the dances such that they are all able to do *something* that is the same and whether that very something is what their ancestors were selected for doing.
>
> (...) when [bee dances] function in the same way that has accounted for the natural selection of their producers and of their answering mechanisms in other workers, they always do exactly the same things. They produce a direction of flight that is a function (mathematical sense) of certain aspects of their form. (Millikan,1998, p. 203-4)[17]

In a nutshell, according to this interpretation bee dances have derived functions not because the dance-producing-mechanism passes them on, but because bee dances form a reproductively established family that acquired functions on their own (Millikan, 2002, p. 130; Sinclair, forthcoming; Martinez, 2010). The idea is that dances (not any particular dance, but dances *as such*) have been selected for because they had these

17 Another passage pointing in the same direction: [It may seem that a particular bee dance has no 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 (...) we 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 that 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 that respect, all bee dances of the same bee species have exactly the same proper function (Millikan, 2002, p.130. Quoted in Martinez, 2010)

further effects. Dances form a reproductively established family (like hearts or other mechanisms) and hence they acquire functions like any other device with a selective story. In other words, bee dances have derived functions *in virtue of* the fact that they belong to the kind *bee dance*, and bee dances have been selected for because they have certain effects on the consumer.

I claimed earlier that functions are most naturally attributed to traits rather than states (see 2.2.3) but, nonetheless, it is also common to attribute functions to states that are products of other functional states. I have already noted that hearts have functions not in virtue of being the products of genes, but in virtue of its own selective story. Similarlty, the heart's function is to pump blood so that it circulates through the blood vessels, but we can also claim that a function of the circulating blood is to bring nutrients to the cells (among others). This is a function of the circulating blood because that seems to be an effect that (partially) explains why blood circulates within organisms. Crucially, this is a function of the circulating blood, i.e. of the product of the heart's performance, but it does not have this function in virtue of a transfer of normativity, but because of its own selective story. So there seems to be a standard way of attributing functions to certain states that are the products of functional devices. In that sense, the idea that there are derived functions (functions of representations and other products of functional devices) seems to be in accordance with ETIOLOGICAL FUNCTION and with much intuitive talk about functions.

However, (and here is where, I think, I am departing from Millikan) if bee dances have functions in virtue of belonging to a type (bee dance) that has been selected for this effect, then there is no clear sense in which derived functions are *derived*. Devices with derived functions acquire those functions from their own selection process, so they seem to have standard direct and relational functions. In the same way that lungs do not derive their functions from the genes that produce them, bee dances do not derive their functions from the mechanisms producing them.[18]

Hence, the only interpretation that makes derived functions compatible with the standard teleosemantic framework entails that derived functions just are direct and relational functions. 'Derived function' refers to the function of particular items that are produced by other functional items, which acquire their functions in exactly the same way any other trait does. So I think nothing justifies talk of an additional sort of functions, much less they deserve being called 'derived'.

Therefore, the functions attributed in 4 and 5 above are essentially the same kind of functions as 1 and 2. Thus, there are only two kinds of functions: relational and non-relational. Talk of adapted relational functions, derived adapted relational functions and so on turns out to be a mere notational variant of the two main categories we have been working with so far. So, in that respect, these distinctions are not

18 Here I think I disagree with Millikan's own view, because she has often claimed that some functions are *literally derived* in the sense of passed on to the products by their producers. Additional evidence for that interpretation: "I was not careful to distinguish in that particular passage between proper functions of adapted devices that are relational but not derived from the producer's particular adaptor (for example, the function common to all bee dances) and *those that are not relational and are derived from the producer's particular adaptor'* (Millikan, 1998, p. 204; emphasis added). If some functions are literally derived, she has to explain how this transference of function between producer and product takes place and why it does not contradict the key insights of the teleosemantic framework.

offering any new tool for dealing with more complex representations. Nothing is lost if we dispense with them.

Nevertheless, I think there is something important this discussion has revealed: the products of representational devices can also have relational and non-relational functions. This idea leads to another question: are the functions of representations (rather than the functions of mechanisms) relevant in establishing their content? If so, even though they do not constitute a different kind of function, they might help us to solve the problems pointed out at the beginning.

3.2.6.4 *Do the functions of representations play a role in content determination?*

In this section I will argue that, even if we accept that representations have functions, they can not play any role in content determination.[19] I have three main reasons for endorsing this view.

First, according to Third teleosemantics and Millikanian teleosemantics, what a representation is supposed to map onto is determined by the relational function of the system that produces the representation (which in turn is determined by the functions of the consumer system). So far in this dissertation I have presented a theory of content for simple representations, so prima facie it is not clear what role adapted and derived functions could play in the picture. Whether specific dances have derived functions (like helping the hive to survive, and so forth) may be an important question on its own, but seems to be orthogonal to the theory of content. If our task is to determine what is the content of a representation, this issue needs to be settled before we attribute any derived functions to the representation.

Secondly, if what we said in the last section was right, the function of representations derives from an effect that all representations *qua* particular kinds of representations have in common; in the case of the pigment-rearranging device, all skin colors have the derived function of camouflaging the chameleon in virtue of belonging to the REF *chameleon's skin color*; bee dances have the derived function of leading other bees to nectar-gathering in virtue of belonging to the REF *bee dance.* I argued that there might be some motivations for thinking that these direct (derived) functions exist.

19 Unfortunately, Millikan is not very clear as to whether derived and adapted functions play a role in content determination (Millikan's 'direct functions' mostly refer to what I have been calling 'non-relational functions'):

> Notice that what would intuitively take to be what is represented in the case of a maladapted bee dance hangs upon the dance direct proper function, (...). It does not hang upon the derived proper function of the dance itself (...). In Part II I will argue that the most dominant notion of what is signed by signs is derived by reference to direct proper functions of the these signs themselves (...). It is not derived by reference to adapted function of the sign's producing devices. (Millikan 1984, p. 43)

The first sentence claims that content is determined by the direct proper function of the dance itself. But, according to Millikan, the proper functions of particular dances are derived from their producing devices, so if bee dances have direct functions at all, they are derived functions, in Millikan's sense. As a result, the first sentence seems to contradict the second sentence.
As I argued in the last section, I think that the direct function of a particular dance derives from something that it has in common *qua* dance with its ancestors. For instance, the direct function of bee dance nr. 879 is to lead bees to nectar. However, I will argue it is hard to see how that function can determine what a dance represents. I think the teleosemantic framework requires that what a sign represents depends exclusively on the functions of the systems that produce and consume it, not on any function of the sign itself.

Now, remember that there are two kinds of derived functions: Non-relational (*bring other bees to nectar*) and relational (*bring other bees to nectar at a particular location f(x)*. It is not hard to see that neither of these functions can play any role in determining the content of a representation.

1. On the one hand, concerning non-relational derived functions, (1) all dances have the same non-relational derived function and (2) this function does not involve any 'mapping function' upon external states of affairs, but it is merely an effect that helps to explain why this kind of dances have been selected for. So it is hard to see how the fact that a particular representation has a direct derived function (like *bring other bees to nectar*) can be of any relevance for content.

2. On the other, derived relational functions do have mapping functions, but the value of these functions always coincides with the value of the mapping function that results from the relational function of the device that produces the dance. In other words, the relational derived function of a bee dance D is *bring other bees to nectar at f(x)* if, and only if, the relational function of the mechanism M that produces D is *produce a bee dance when there is nectar at f(x)*, where both variables will always get the same value. So, if derived relational functions were relevant for content determination, it seems that the dance's derived relational proper function would produce the bee to represent nectar at L only if the dance already represents location at L (in virtue of being the product of a mechanism with a certain mapping function). So the relational function of bee dances does not contribute in any way to the content of the representation. It is redundant. Therefore, neither direct nor relational derived function can contribute in any interesting way to the content of the representation.

The third reason for thinking that derived functions are irrelevant for content determination is that if the content of bee dances depended on the functions of particular dances, then it would not be clear why we need the whole sender-receiver framework in the first place. If bee dances *as such* are selected for (as our interpretation of derived functions suggests) and their content is somehow determined by this selection process, the whole theory of producer and consumer systems is not doing any real work in naturalizing content.[20] Consequently, whether particular representations or states have further functions is something that cannot affect our ascription of content to mental states.

Let us take stock. This discussion aimed at showing that there is a particular phenomenon Millikan's notions of adapted and derived proper function was trying to capture. Nonetheless, we found that the so called 'derived proper functions' and 'adapted proper functions' are nothing more than direct and relational proper functions of certain states that happen to be the products of other devices with functions. Derived proper functions are not a new category of functions and, more importantly, require selection processes in exactly the same way other

20 Of course, this argument is only compelling for those that accept some form of sender-receiver model (which I defended in 2.2.3). In that respect, Millikan's emphasis on the existence of systems that produce and consume representations suggests that she would agree that this structure is relevant in content determination.

functions need them. Furthermore, they cannot play any role in content determination within a teleosemantic theory of content.

Thus, since adapted and derived functions do not provide any new tool for analyzing representations, I suggest to dispense with these notions and talk instead of direct and relational functions *tout court*, as we have been doing so far.

Here concludes the first part of this chapter, whose goal was to consider some ways in which the first versions of teleosemantics had to be improved. Now that we have a more precise and sophisticated toolkit, let us consider some old and recent objections and see whether Third teleosemantics can satisfactorily deal with them.

3.3 objections

After considering some problems and limitations of the theory, in the remainder of this chapter I would like to discuss some objections and alternative teleosemantic theories: (1) Neander's Producer-based Account, (2) the alleged circularity of teleosemantics, (3) Shea's Infotelsemantics, (4) the possibility of uncooperative representational systems, and (5) the counterexample of Swampmen. Of course, there are many other objections and alternative accounts to teleosemantics, but I have decided to address this set of questions, either because they are crucial issues in the debate (this is the case of 1 and 5), or because I think they raise certain issues that nobody has satisfactorily addressed in the literature (that happens with 2, 3 and 4). In what follows, I will argue that Third Teleosemantics can provide adequate replies to all these questions.

3.3.1 *Neander's Producer-based Account (1995, 2006)*

Neander has put forward a theory of mental content that differs from the rest of accounts in important respects. Like other teleosemanticists, she thinks representations are the products of certain biological mechanisms whose function is to produce states that are supposed to correlate with the presence of certain states of affairs. However, her solution to the Indeterminacy Problem greatly differs from others' and has given raise to an important and original theory of mental representation.[21]

3.3.1.1 *Functional analysis*

Neander's original proposal discusses the very foundational claims we have been concerned with in these first chapters.

First, she argues that one and the same trait can have many different effects. For example, a gene in antelopes alters the structure of hemoglobin, which causes higher oxygen uptake, which in turn allows the antelope to survive to a higher ground. Neander points out there is a by-relation between these effects: genes allow the antelope to survive to a higher ground *by* increasing the antelope's oxygen uptake, and the latter is achieved *by* altering the structure of hemoglobin. But, what is the function of the gene? She remarks that, on the etiological notion of function, all of them are functions of the genes, since all of

21 Let me point out that, while I am writing these lines, Neander is working on a new teleosemantic approach, which combines her teleosemantic account with some ideas of informational theories.

them are effects (some more proximal, others more distal) that explain why these genes have been selected for. Nonetheless, Neander points out that there is such thing as a trait's *primary function*: the primary function of a trait is identified with the function that corresponds to the lower level at which the trait is an unanalysed component. In the case of the antelope's gene, it seems that the function that corresponds to the lower level of analysis is altering the structure of hemoglobin. So this is its primary function. Why the lower level of analysis has this privileged status is something I will consider in short. The key point is that, on Neander's view, the relevant function for content determination is the primary function, which is defined in relation to the 'lower level at which the trait in question is an unanalysed component of the functional analysis' (Neander, 1995, p.129).

How does this theory on functions bear on the context of a theory of representations? Neander assumes that once the relevant notion of function is established, its application to the case of a representation is straightforward. Think again about frogs. She argues that, if we focus on the lower level of analysis, the function of the prey-mechanism is to detect black moving shadows and not flies or nutritious things; this is so because the frog detects flies *by* detecting black shadows and detects nutritious things *by* detecting black moving shadows. In the same way that in the case of hemoglobin we take the most proximal effect (the effect at the bottom of the by-relation), content is determined by the function at the lower level of analysis. The result is that the content of a representation is usually identified with its proximal cause.

3.3.1.2 *Assessing Neander's Account*

There are some issues in Neander's account that need to be spelled out in detail in order to rightly evaluate her proposal. For instance, while she is very concerned with the description of the relevant notion of function, it is not entirely clear how this notion of function is used in a theory of representation and content. In particular, it is obscure whether she holds that the content of representation is determined by the function of the representation or the function of the mechanism that produces representations.[22] In that respect, concerning the famous frog example, Neander claims that the lower level of description *of the detection device* is the one in which the frog's mechanism detects black moving things, so she concludes that the *function of the representation* is to represent the presence of a black moving thing. As I argued earlier, I think there are good reasons for thinking that representational content is not determined by the function of the representation, but by the function of the mechanism that produces representations. Indeed, it is hard to make sense of the idea that the function of a state determines its meaning. So, in order not to be unfair to her proposal, I will interpret the expressions that appeal to the 'function of a representation' as a loose way of talking.

Hence, in order to apply Neander's recipe, we should consider the lower level at which the representational *system* is an unanalysed component. Let me try to formulate her view more precisely:

Neander's Theory

1. Any trait t has a set of effects E.

22 Let me say here that Neander is not alone here; most teleosemanticists fail to adequately draw this distinction.

2. There is a by-relation between the different effects $e_1, e_2, e_3, ..e_n$ that compose E. E is a linear order.
3. The primary function of a trait t is e_1, that is, the lower bound of the linear order E.
4. The content of a representation M is determined by the primary function (e_1) of the trait t producing M.

Now, I think the main problem of this theory is that claims 1 to 4 fail to determine an univoque content for M. The reason is that when we focus on the cases that involve representational systems, there is no way of determining the primary function of a system (condition 3). That happens because the alternative effects that give rise to the indeterminacy problem do not form a linear order in the way condition 2 requires. So Neander's Theory fails to provide a single and determinate content. Let me explain.

For Neander's proposal to work, there has to be a linear order among the effects of t. But the by-relation determines a linear order E only if it is an asymmetric relation. Indeed, very often the by-relation is asymmetric. Most of the time, this asymmetry of the by-relation is grounded in some causal and temporal relations. For instance, I can break a window by throwing a ball, but I cannot throw a ball *by* breaking a window.

However, there is sometimes a by-relation between different states that is symmetric. For instance, I break a glass *by* breaking the window and I break the window *by* breaking a glass; similarly, I am calling the President of the USA *by* calling Mr. Obama and I am calling Mr. Obama *by* calling the President. In these situations the by-relation is symmetric, and for this reason it does not establish a linear order. As a result, there is no single lower bound and Neander's proposal cannot be employed.

Now, consider the case of frogs. An effect of the mental state is to produce M when there is a black shadow around, another effect is to produce M when there is a fly and still another to produce M when there is a nutritious thing. But is the relation between these effects symmetric or asymmetric? Well, the frog produces an M when there is a fly *by* producing an M when there is a nutritious thing; and it is also reasonable to claim that the frog produces an M when there is a nutritious thing *by* producing and M when there is a fly. Both seem to be legitimate descriptions of what is going on. Similarly, the frog produces M when there is a fly *by* producing M when there is a black thing, but also produces M when there is a black thing by producing an M when there is a fly. So there is no effect e that can be considered the lower bound of the analysis. Indeterminacy threatens.

Neander may complain that some of the last claims are false. For instance, she might deny that the frog produces M when there is a black thing *by* producing an M when there is a fly. But why so? After all, both black things and flies cause the mental state to go on. Moreover, the presence of M increases the probability of there being a black shadow and also the probability of there being a fly and a nutritious thing.[23]

23 Of course, she could argue that the frog produces an M when there is a fly *by* producing an M when there is a back thing around (and not vice versa) because it correlates better with instantiations of *blackness* than with instantiations of *flyhood*. But, if she took this option, all the work would be done by this notion of *correlation*. Accordingly, Neander's proposal would be classified as a version of Relative Indication, which was discussed and rejected in 1.2.3.3.

Crucially, I think that the only way one can deny this claim is by assuming that M represents black shadows. Certainly, if M represents black shadows, then the frog produces M when there is a fly *by* producing M when there is a back shadow, and it is false that the frog produces M when there is a black shadow *by* producing M when there is a fly. But, similarly, if one assumes that M represents flies, it is true that the frog produces M when there is a black shadow *by* producing an M when there is a fly and false that the frog produces M when there is a fly *by* producing an M when there is a black shadow. The result, then, is the following: the by-relation is only asymmetric (and hence NEANDER'S THEORY applies) only if we already assume that the content of the representation is such and such. So Neander's account either gives a highly indeterminate content or begs the question against her opponents.

Let me put the same idea in a different way: Is the mechanism's *primary* function to produce M when there is a fly around or when there is a black thing around? I hope the difficulty in answering this question is obvious: we do not know which is the primary function in that case, because this is is precisely what we are trying to settle. If we knew which are the fundamental relata that ground the by-relation, we would have a single content; but there is no way of finding out a primary function without previously assuming that the state has a certain content. So Neander's account of primary functions and levels of analysis fails to determine a content for the state.

Consequently, when she claims that the primary function is to produce M when there is a black shadow, which is a determinate content, she is begging the question against other candidates. Merely focusing on primary functions and levels of analysis does not yield this result.[24]

In any event, Neander thinks that what the 'lower level of analysis' yields is a representation of a moving black thing rather than fly. Let us grant that this is the result of assuming this perspective; we can still reasonably ask the following question: why does the lower level of analysis have this privileged status? Why should we assume that content is determined in that way? Neander offers four arguments; I think three of them are unsatisfactory. Nonetheless, I will argue that the last one points in the right direction, although it fails to favor her account over mine.

INFORMATIVENESS The first reason she gives is that, if content is determined by the lowest level of analysis, then it makes talk of malfunctioning more informative. If we assume that the function of the frog's mechanism is to detect black moving things and we know that

24 Jacob (2000, p. 19, emphasis added) offers a similar proposal: "In a word, instantiations of property G [flyhood], not F [black moving things], help explain the proliferation of creatures with mechanism M. But again, explaining the proliferation of M is not fixing the content of M-states. *Given the creature's sensory limitations, it is only by means of their representation of F-instantiations that such creatures can tell when to act appropriately, i.e. when G is instantiated*". He merely assumes that content is determined by sensory limitations -but this is precisely the question we are trying to settle.
Let me add that Jacob's main reason for favoring a representation of F over G is surely mistaken. He thinks it follows from the following principle, that he calls 'Nomic Correlation Principle': unless its tokenings are nomically correlated with instantiations of property F, sensory mechanism M cannot represent property F'. If 'nomical correlation' means something like statistical correlation (e.g. $P(\mathtt{fly}|M) > P(\mathtt{fly})$), mental states M have a nomical relation with black moving shadows but also with flies. If it means something stronger, then it is not clear that this Principle is right, for the reasons adduced when discussing indication theories in 1.2.3.

in a certain occasion the mechanism failed to fulfill its function, we know that he has not reacted to a black moving thing. If we thought the function is (say) to catch a fly, knowing that it has malfunctioned does not tell us whether the problem laid in the detection of black moving things, the absence of a fly, the snapping mechanism, etc..

The first problem with this argument is that assuming that the represented state of affairs is determined by the lowest level of analysis makes malfunctioning more informative *at the cost of making proper functioning much less informative*. Being told that the system works properly is much more informative if the system is supposed to catch flies than if it is supposed to detect black moving things. So, Neander's account renders malfunctioning more informative, and other accounts make proper functioning more informative. Prima facie it is not obvious why we should prefer one option rather than the other. So I doubt that anything is gained in informativeness by adopting Neander's view.

But, more importantly, a problem with this argument is that it is not clear that attributions of content should depend on how informative they are for us. If we all accept that *having such and such representational content* is a real and objective property of certain entities, then it is not obvious why the informativeness of a certain attributions should affect our predictions about content (for a discussion on the relation between content and informativeness, see 4.1.3).

SPECIFICITY The second argument Neander puts forward for favoring the lower level of analysis is that the lower the level of analysis is, the most specific is the function we are picking up. The specific function of a trait is supposed to be the function that a trait can perform independently of the others (see also Papineau, 1998; for a discussion, 5.2.3.1). The idea is that heart's pumping blood is the *specific* function of hearts, which differs from other functions in the fact that hearts can perform them only in conjunction with other traits. According to Neander, all of the heart's effects are caused by the heart plus other parts of the human body, except pumping blood. She suggests that the specific function of a trait, which the lowest level of analysis helps to bring forward, is something the trait does alone. This is one reason for favoring it as the most adequate level of analysis.

Now, one may wonder whether traits have any *specific* function in that sense. Hearts pump blood only when they are supplied with blood, they are appropriately connected to other parts, they are sustained by the organism,... In fact, that seems to be the key insight of the Organizational Account of Function discussed in the previous chapter, according to which any function requires the activity of a diverse and differentiated self-maintained system (Mossio and Moreno, 2010; see 2.1.2.4). Similarly, for the frog to detect black moving things, the visual system has to operate in the appropriate way, which includes the proper functioning of the retina, the optic nerve, etc,... Consequently, I doubt there is any specific function of traits in the intended sense. There is no effect of a trait that does not depend in one way or other upon other parts of the body.[25]

25 Compare with Papineau (1998, p. 15, emphasis added): 'So if we want to identify effects which it is the function of the desire to produce, we need to go far enough along the concertina to reach results *which do not depend on which beliefs the desire happens to be interacting with*.'

PROXIMAL CAUSES I think that what Neander had in mind when formulating NEANDER'S THEORY is that the lowest level of analysis is the one in which we appeal to the most proximal cause of a given mental state. Furthermore, she thinks that this is the right result because (at least, cognitively unsophisticated organisms) represent such features as black shadows, shapes and motion.

However, if the claim is just that content is determined by most proximal cause, there is an imporant problem lurking ahead: the most proximal cause of the frog's representation (or representational system) is not the presence of an external black moving thing, but the presence of light impinging the retina, activation of the cells in the retina or perhaps the activation of cells in the optic nerve. Indeed, since this discussion is about the conditions for *any* representational system, this account would entail that *all* representations can only represent proximal causes. This is surely an unsatisfactory result 2.3.3.1. As she writes:

> For any description I give that speaks of distal objects (e.g., small dark moving things) can be trumped by one that speaks only of a proximal object (i.e., a retinal pattern of a particular kind). It is, after all, by responding to a retinal pattern of a particular kind that the frog responds to small dark moving things. (Neander, 1995, p.136)

Now, Neander admits that her account fails to solve this problem (what she calls the 'distality problem') and she replies as follows:

> Now, it's true that I haven't provided a principled answer to the distality problem, but I haven't precluded one either, and, in fairness, it was not the problem being tackled. (Neander, 1995, p.136)

Still, her theory was supposed to provide a satisfactory theory of content and we saw that any account with this aim must satisfy a minimum set of desiderata, one of which is solving the adequacy problem. So not having precluded a solution is an insufficient reply.

COGNITIVE SCIENCE While in her main work she relies on the three arguments just discussed, in a more recent paper Neander (2006) justifies her view in a different way, which I think is more promising (see also Schulte, 2012). In Neander (2006) she makes clear that her main reason for claiming that the content of the frog's mental state is to represent black moving things is that this attribution is in accordance with cognitive science (in particular, with neuroethology). When cognitive scientists describe the content of mental states, specially in early perceptual processing, they often attribute representations of, say, vertical lines, round figures or black items. Since this is the kind of representations that cognitive scientists attribute, Neander thinks that the claim that organisms such as frogs or salamanders represent distal objects is in tension with standard scientific practice. I think this is a powerful argument (see also Jacob, 1997, 2000; Jacob and Jeannerod, 2003) and one that cannot be dismissed without serious consideration.[26]

In chapter 4 (see 4.1.4.2) I will extensivley argue that the teleosemantic proposal of the sort I suggest can accommodate scientific practice.

26 On the other hand, since this is precisely the argument employed by supporters of RELATIVE INDICATION, it lends some support to the idea that Neander was actually basing her account on some sort of correlation, as I discussed in 3.3.1.2.

Therefore, whereas I completely agree with Neander that a teleosemantic account should be sensible to actual practice in neuroscience, that desideratum fails to support Neander's approach over my own teleosemantic view. This point will be argued for in the next chapter.

In conclusion, Neander's proposal is insufficiently motivated and has significant difficulties. So I think Third Teleosemantics provides a much better framework for a naturalization of representation and content.

Let me now move to the objections that some people have raised against the sort of teleosemantic framework that I have defended.

3.3.2 *Circularity*

In a recent paper, Shea (2007) has raised a sensible objection to mainstream teleosemantics and an alternative theory with a broad teleosemantic inspiration. In this section, I would like to consider his objection to teleosemantics and show that Third Teleosemantics is preferable over his own proposal.

Shea's objection is based on one of Godfrey-Smith's (1996) criticisms. Originally, Godfrey-Smith's put the problem as follows:

> For correspondence to have a real role in the production and explanation of success, it must be conceptually distinct from the fact of success. Success-linked theories threaten this independence. (Godfrey-Smith, 1996, p . 192)

Let me carefully spell out the ideas contained in this succinct quote. In general, an adequate way of explaining a successful behavior is by appealing to the fact that a subject had true representations. The fact that John had a true belief about there being a beer in the fridge explains why he went to the fridge, opened it, took a beer and thereby satisfied his desire for beer. That looks like a satisfactory (even if partial) explanation of how he managed to achieve his goal.

Now, the worry Godfrey-Smith is pointing out is that it seems that teleosemantics has the consequence that attributions of true representations do not provide any substantial explanation of why a certain behavior was successful. Shea (2007) makes a parallelism with Dr. Pangloss in Moliere's *Le malade imaginaire.* When asked about why a certain pill causes people to immediately fall asleep, Dr. Pangloss mentions the fact that it has a dormitive virtue, which is another way of saying that it has the disposition to cause people fall asleep. But, of course, an explanation in terms of dormitive virtue does not look like an explanation at all. Similarly, Godfrey-Smith and Shea argued that, according to teleosemantics (and, more generally, according to success semantics), the fact that a representation has an accurate content is explained by the fact that this condition prompted successful behaviors; but then, as in the dormitive virtue example, it seems that an explanation of a given successful behavior in terms of having a true belief is not explanatory at all.[27]

27 It is worth mentioning that this is an objection not only against Millikan's version of teleosemantics but also against other versions, such as Dretske's: "Once C is recruited as a cause of M - and recruited as a cause of M because of what indicates about F- C acquires, thereby, the function of indicating F. Hence, C comes to represent F. (...) What you believe is relevant to what you do because beliefs are precisely those internal structures that have acquired control over output, and hence become relevant to explanation of system

The problem can be traced back to success semantics. Success semantics defines the content C of a belief R as that condition that accounts for the success of the behaviors prompted by R. Now, since R is defined by appealing to success, it seems that the presence of R cannot explain why the behavior was successful. One cannot explain the success of a behavior B by appealing to the content of a true representation R, and then explain the fact that R has the content it has in virtue of causing the success of B. Indeed, Shea (2007, p. 430) argues this problem is a particular version of a general worry about functionalism: if one defines a state R by appealing to a set of effects of R, one cannot explain these effects by appealing to R. If one defines *being in pain* as the state that causes certain avoidance behavior, one cannot in turn explain avoidance behavior by appealing to pain.

Now, Shea (2007, p. 413) admits that if the goal is to explain the successful behavior of a particular subject, mentioning the fact that he has a true representation R with content C has some explanatory import. For instance, it excludes the possibility of this behavior succeeding due to another agent or due to mere luck. However, he argues that this is only *thinly* explanatory: it merely subsumes the behavior of a particular agent under a regularity.[28] It does nothing to explain why having a true representation led to success.

More precisely, Godfrey-Smith and Shea think that teleosemantics faces this problem of circularity:

1. According to Teleosemantics, having a true belief is explained by appealing to successful behavior prompted by these states.

2. If true belief is explained by appealing to successful behavior, true belief cannot explain success.

3. According to Teleosemantics, true belief cannot explain success.

4. True representations are relevant to explaining the success of behavior they cause. (Having true beliefs is 'fuel for success')

   ∴ Therefore, Teleosemantics is false.

I have already justified premises 1, 2, and 3. Let me shortly comment on premise 4.

Premise 4 states that having a true representation is a 'fuel for success' (using Godfrey-Smith's expression). However, the idea that semantic properties are not really explanatory is also an old one (Field, 1978). So one option a teleosemanticists could take is simply to deny 4 and admit that, according to his theory, attributions of true content are not explanatory. What is wrong with this straightforward reply?

There are various reasons that advise not to follow this line of response. First of all, there would be a strong tension between the efforts that teleosemantics puts in accounting for representational content and their claim that semantic properties are not really explanatory. Secondly, as many people have pointed out, the idea that semantic properties are

behavior, in virtue of what they, when performing satisfactorily, indicate about external conditions." (Dretske, 1988, p. 85)

28 Millikan (2007, p. 438) agrees with Shea that, when explaining the successful behavior of a particular organism, appealing to its having true content excludes it being caused by other things. However, she replies that this kind of explanation is not as thin as Shea suggests. It is hard to know how to settle this dispute. In any event, for the sake of the argument, I will to grant that these explanations are too thin for being substantive.

explanatory is extremely plausible in itself (it is part of what philosophers call 'folk psychology'). Finally, teleosemantics could be accused of offering an ad hoc solution to the circularity objection. These reasons strongly suggest that teleosemantics should hold premise 4 and look for another kind of solution.

Shea thinks that this argument shows that THIRD TELEOSEMANTICS needs to be supplemented with an informational condition. That is the reason he develops his Infotel-semantic account. He claims that once we add the condition that a true representation must carry information about the state it represents, explanations in terms of true representations become much more explanatory. Let us now describe and discuss this 'Infotel-semantic theory', before directly addressing the circularity objection to THIRD TELEOSEMANTICS.

3.3.2.1 *Infotel-semantics*

In order to flesh out Shea's proposal in more detail, we need two things: first, we have to define the relevant notion of information and, secondly, we need to specify how it can be included into a teleosemantic account.

On the one hand, Shea defines a notion of information that resembles very much our WEAK INDICATION (see 1.2.3.2) and Millikan's 'locally recurrent natural information'. According to Shea:

SHEA INFORMATION R carries the correlational information that condition C obtains iff for a common natural reason within some spatio-temporal domain D, chance (C | R is tokened) > chance (C | R is not tokened)

Basically, a state R carries correlational information about C iff R weakly correlates with C and there is some common natural reason (often some kind of causal relation) that underpins these probabilities (see also Martinez, 2010).

Notice that, in this sense, information does not require the existence of a natural law between R and C, since the kind of correlation required by SHEA INFORMATION is relative to a certain spatio-temporal domain (Cf. Fodor, 1990). This is a reasonable assumption given that, very often, the correlation between signs and what they signify obtain only in a very restricted domain (see 1.2.3).

The idea, hence, is to supplement the standard teleosemantic account with an informational condition, which be able to account for the explanatory import of intentional explanations. Thus, Shea develops what he calls 'Infotel-semantics', according to which (Shea, 2007, p. 418-9):

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

(a) 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 (a) to (d);

(b) Rs carry the correlational information that condition C obtains.

(c) An evolutionary explanation of the current existence of the representing system adverts to Rs having carried information about C; and

(d) C is the evolutionary success condition, specific to Rs, of the behavior of the consumer prompted by Rs.

Conditions (a), (c) and partially (c) are intended to capture the standard teleosemantic claims, which I set up in Third Teleosemantics.[29] The key innovation is condition (b) and part of (c), which suplements standard teleosemantics with an informational input condition.

In a nutshell, Shea's suggestion is that by introducing an appeal to information, which is not defined in terms of a success condition but in terms of mere correlation, the circularity problem is immediately solved. Since the fact that a state has a certain content is not only explained in terms of success conditions but also in terms of carrying certain information (which, in turn is defined in terms of correlation), the claim that a subject truly believes P provides a substantive explanation of its success. The informational bit is supposed to make a crucial contribution; it renders a circular explanation into a non-circular one.

Interestingly enough, this idea was already pointed out by Godfrey-Smith (1996, p.184):

> Dretske's theory, because of its residual appeal to indication, does have the potential to preserve more of the idea that truth is a fuel for success than the other theories discussed in this section.

In the next section I will address Shea's Infotel-semantics. I will present two objections that show that Infotel-semantics is not better (and probably worse) than Third Teleosemantics. Afterwards, I will argue why, despite the plausibility of the analogy with Dr. Pangloss, Third teleosemantics is really explanatory.

3.3.2.2 *Problems with Infotel-semantics*

It is worth mentioning that the idea of combining teleosemantics with some sort of informational theory is not a new one (e.g. Dretske,1995; Prinz, 2002, ch 9; Neander, 2013), although the details of Shea's proposal and his motivations are original. Nonetheless, I will argue that there are two aspects specific of Infotel-semantics that pose serious difficulties.

is information sufficient? First of all, if we grant for the sake of the argument that teleosemantics suffers from circularity, it is not clear that the notion of weak correlation can do the job Shea wants it to do. In particular, if teleosemantics renders explanations in terms of true representations circular, it could be argued that Infotel-semantics falls prey to the same problem. Let me elaborate on that point.

Notice that, in general, carrying information (in the sense of Information) is very cheap. Any given representation carries a huge amount of information about a wide range of entities. For instance, Fin Whales (*Balaenoptera physalus*) perform low frequency calls that correlate with the breeding season, with the presence of a whale male (only males perform these calls), with seasonal migration, with the absence of sea ice concentrations in the area and with many other facts (Watkins et al, 1987; Croll et al, 2002; Sirovic et al. 2004). Merely increasing the probability of another event occurring in a certain domain due to some natural reason is not very hard to satisfy.

29 Interestingly enough, notice that it resembles very much First Teleosemantics, i.e. the simplest version of teleosemantics.

Of course, in contrast to other accounts (e.g. Dretske, 1981; Neander, 2013) the fact that representations carry information about many states of affairs does not imply that Shea's theory have problems with the indeterminacy of content, because he also adopts the main insights of a teleosemantic account (specially the appeal to consumer systems). The teleosemantic component of Infotel-semantics has the consequence that most of the entities a state correlates with do not figure in its content. However, the fact that any state correlates with a great amount of features threatens his solution to the circularity problem, because the weak correlation between representation and representatum is supposed to make all the explanatory work that teleosemantics by itself is unable to make. While (a) to (d) are required for a state to be endowed with representational content, (b) is the condition that is supposed to account for the explanatory import of teleosemantic explanations. But can a weak correlation by itself turn an unexplanatory attribution into a fully explanatory one? That probably requires too much from such a notion. A weak correlation seems to be too easy to satisfy to ground the explanatory import that (according to Shea) is missing in teleosemantics. Hence, if Shea is right and teleosemantics only provides thin explanations of behavior, infotel-semantics probably falls prey to the same problem.[30]

There is a different way of spelling out the same worry. States carry correlational information in virtue of the fact that certain statistics hold (due to some underlying reason), so if true representations carry correlational information about a state C, so do false representations. That is, both true and false representations carry exactly the same information. So why should we think that the fact that a state has information explains the *success* of a particular behavior? Perhaps carrying information can explain why I act as I do, but the circularity problem concerned *successful* behavior. Shea's argument is not intended to show that teleosemantics renders explanations of behavior in terms of content unexplanatory. The objection had to do with the explanation of success; and if true and false representation both carry information, it is unclear to what extent carrying information can explain success.[31]

Therefore, there are some reasons for thinking that the notion of correlational information included in Infotel-semantics might be too weak to supplement the explanatory value that is allegedly missing in teleosemantics.

PRODUCTIVE REPRESENTATIONS Secondly, by including the condition that in order for R to represent S R has to correlate with S, Infotel-semantics loses any chance of providing an account of the

30 Notice that condition (c) of Infotel-semantics ('An evolutionary explanation of the current existence of the representing system adverts to Rs having carried information about C') cannot be appealed to in order to solve this issue. It is the fact that a state carried information (condition b of Infotel-semantics) that is supposed to solve the circularity problem, not the fact that information played some role in the evolution of the mechanism. If Shea relied on condition (c) in order to increase the explanatory value of the notion of information, he would incur in exactly the same problem as teleosemantics: the explanatory role of information would depend on its figuring in an explanation of the selection of the representational mechanism, that is, it would depend on information having played a role in accounting for successul behavior.

31 In that respect, there is a key difference, for instance, between an appeal to information and an appeal to causal relations. The fact that my representation R has been caused by C can (partially) explain why my behavior was successful. After all, C can cause R only if C holds. However, the fact that R carries information about C does not entail that C holds, so it is hard to see why it should contribute to an explanation of success.

productivity of some representational systems. As I pointed out in 3.2.4, there might be many contentful representations that are tokened just once (e.g. a particular bee waggle dance indicating nectar at 235m in such and such direction or the missing shade of blue). Since they are just produced once, they lack the statistical correlation that is required for satisfying Shea Information so they are rendered contentless by Infotel-semantics. Consequently, Infotel-semantics cannot accommodate the capacity of many organisms of producing new contentful representations.

Shea could reply that even representations that are tokened just once can carry information, because his weak notion can be satisfied by two states that have correlated only once in a very specific environment. Unfortunately, this reply will not do, for the simple reason that a contentful representation can be tokened just once and furthermore be *false*. Suppose it is the first time a bee performs the waggle dance n.876, which indicates *nectar at 235m in such and such direction* and suppose it turns out to be false (there is not nectar at the position it signals). Accordingly, there is no correlation between the dance and the source of nectar, so the state does not satisfy condition (b) and (c) of Infotel-semantics. Consequently, Infotel-semantics entails this dance is a contentless state.

Notice that this is a general result: any representational mechanism that exhibits productivity will give rise to some states that seem to be perfectly well formed and contentful, but which would be rendered contentless by Infotel-semantics. By adding an informational input condition, this approach is unable to account for productivity.[32]

Consequently, Shea's approach loses much of its plausibility when we focus on complex representations. Indeed, Infotel-semantics would probably lead us to a 'splitting account' of representation, according to which the representations of simple organisms are different in kind from the representations of more complex organisms (see Shea, 2007, p. 419; 2013). Some people might be happy with that result (Burge, 2010; Rescorla, 2013), but most teleosemanticist would strongly disagree with it (Millikan, 1984; Neander, 2013; Papineau, 1993; Price, 2001).

I think both problems can be overcome by Third Teleosemantics. On the one hand, teleosemantics is not threatened by the second drawback because it is not based on any kind of correlation between representations and representata. On the other, I think Shea's objection can be met by Third teleosemantics, so in fact there is no circularity problem and Infotel-semantics (or any appeal to information) is not required. Let me show why I think the worry of circularity is unfounded.

3.3.2.3 *Solving the Circularity Problem*

First of all, it is important to notice that premises 1 and 2 of Shea's argument can be interpreted in two ways, and each of these interpreta-

32 Shea (2007) claims on footnote 19 that he is restricting his attention to 'representation in simple organisms'. Nevertheless, he explicitly states that he wants to account for the representational abilities of bees, among others, so he is supposed to deal with some productive representational systems. Furthermore, remember that probably most signals are productive in the sense defined here (Millikan, 2004). Think, for instance, about the frog's brain states, which represent something like *there is a fly around now*; it is very implausible that the representational state carries correlational information of that particular time and place.

tions pose a different problem for teleosemantics. Those are the two premises of his main argument:

(1) According to Teleosemantics, having a true belief is explained by appealing to successful behavior.

(2) If true belief is explained by appealing to successful behavior, true belief cannot explain success.

On the one hand, the circularity can concern the fact that a state has a certain content. In other words, Shea could be arguing that according to teleosemantics:

(Circularity Content) Having a *belief about C* is explained by appealing to successful behavior, so the content of a belief cannot in turn explain successful behavior.

On the other, (1) and (2) can be interpreted in a different way. Premise (1) can be read as stating that the circularity concerns *true* beliefs, rather than mere beliefs. The claim would then be:

(Circularity True Content) Having a *true belief about C* is explained by appealing to successful behavior, so the content of a belief being true cannot in turn explain successful behavior.

The supporter of Third Teleosemantics should reply to these two challenges.

CIRCULARITY OF CONTENT Let us focus first on the circularity of Content. There are three aspects that show that content attributions according to teleosemantics are much more explanatory than the ascription of properties of the dormitive-virtue type.[33] I will present them separately, since I think each one probably suffices as a reply to Shea's and Godfrey-Smith's concerns. But if they are put together, the provide strong support for the view that third teleosemantics is perfectly compatible with content attributions being explanatory.

Backward-looking Properties The circularity objection assumes strong similarities between attributions of semantic properties according to

33 Millikan's (2007) main line of response appeals to something like Dretske's distinction between *structuring* and *triggering* causes (Dretske, 1988). Triggering causes explain why an event occurred at certain time while structuring causes account for the process that shaped or structured the process. For instance, if we explain the fact that Mary stood up *then* by saying that the Queen entered the room, we are providing a triggering cause; if instead we mention the fact that standing up is a gesture of respect, we are providing a structuring cause.
Now, Millikan argues that Shea's objection to the explanatory import of attributions of true beliefs is based on the assumption that only triggering causes are explanatory. Accordingly, she replies that teleosemantics does not make belief ascriptions superfluous because belief attributions describe the kind of mechanism that produces action (so they provide a structuring cause), even if they do not mention the triggering cause of the behavior. Teleosemantics offers historical explanations, and history can only explain why there is a connection between a state of the organism and particular behavior B, not why this behavior occurs at a certain time. So this is the sense in which teleosemantics makes content attributions explanatory.
Unfortunately, I think this response in unlikely to succeed. It is not true that Shea only admits triggering causes as explanatory. For instance, he claims that dispositional properties are sometimes explanatory. Indeed, his own notion of correlational information could be classified as a structuring cause (see below). Shea's objection to teleosemantics is that the kind of facts mentioned in the definition of the structuring cause *includes the fact* that it causes this successful behavior. The problem is that the explanandum is an element in the definition of the explanans, so it makes the whole explanation circular.

teleosemantics and dispositional properties such as the property *having dormitive virtue* ('having dormitive virtue' refers to the dispositional property *being disposed to cause people fall asleep*). In particular, Shea claims that both properties include the particular instance they are explaining in the definition of what having the property is. However, there is a crucial difference between attributing a dispositional property like fragility, fitness or dormitive-virtue and attributing a semantic property according to teleosemantics. Fragility is (roughly) the propensity to break under a wide set of circumstances and fitness is usually defined as the propensity to survive and leave a certain amount of offspring (Sober,2002, p. 319). They are, so to speak, *forward-looking* properties. In contrast, according to teleosemantics semantic properties depend on what has already happened, so they are *backward-looking* properties. The attribution of semantic properties exclusively hangs upon (very complex) *past* facts.[34]

Why should we think this apparently minor difference is so important? In that particular case the appeal to history makes a crucial difference, because when defining the fact that R means C we are not including the actual situation. Dispositions are sensitive to what is actually the case. No one has the disposition to cook dodos, because there are no dodos any more (Millikan, 2004). In contrast, Third Teleosemantics does not define content attributions by appealing to what the organism is supposed to do in the present situation. It is, so to speak, blind with respect to the present. And remember that, in general, Godfrey-Smith and Shea accept explanations in terms of dispositions, functions and the like. The key problem of circularity they are pointing out depends on the fact that having a disposition (a function, etc.) is defined by appealing to the success of one's behavior. The circularity problem is to include the particular cases one is trying to explain in the definition of what it is to have a mental state with a certain content. In contarst, if R is defined (constitutively) as the entity that caused b_1, b_2, b_3,... b_n, there is no reason why a token of R cannot explain its successfully causing b_{n+1}.

Let me put the point in a different way. Biologists usually distinguish *being an adaptation* from *being adaptive* (Sober, 1984, p. 120). The adaptiveness of a trait depends on the extent to which a phenotype fits its local ecological niche. Accordingly, saying that organisms with a certain trait survive better in a given environment because this trait is adaptive is only thinly explanatory. This is an explanation of the dormitive-virtue type, because in order to ascribe adaptiveness one is already considering the current situation. In contrast, a trait is an adaptation when it is the result of a process of selection. An adaptation must have been adaptive, but might not be adaptive in the current environment. Tusks are probably adaptations of elephants, but they are not adaptive any more: every year thousands of elephants are being hunted because of the ivory. Now, Third teleosemantics does not claim that representational systems are adaptive; it only entails that they are adaptations. And since the ascription of content does not take the present situation into account, having a certain content can be fully explanatory.

Notice that this backward-looking aspect of the theory is rooted on the notion of etiological function that is essential to teleosemantic theories. On the etiological understanding of function, whether a trait

34 Of course, demonstrative and indexical expressions are exceptions.

has a function depends on the *past activities* of traits of the same type, not on any feature that this particular trait does or has the capacity to do. One virtue of this account is that it can attribute the function F to a trait even if this particular trait is unable to perform F (see 2.1.2). Similarly, it seems that we can satisfactorily explain why a particular heart pumps blood by mentioning the fact that this is its function. The fact that it has a function does not presuppose or entail that it will successfuly pump blood *now*.[35]

This backward-looking aspect of teleosemantics is a fundamental feature of the theory. It shows that the facts in virtue of which teleosemantic theories attribute functions is very different from the facts in virtue of which we attribute dispositional properties to entities. Indeed, this idea was hinted at by Godfrey-Smith (1996, p.182)

> (...) the success used to determine meaning in these theories is past success, and we assume here that we are trying to explain a present episode- an episode occurring after the inner states have acquired a truth-condition. Teleonomic theories do not, strictly speaking, assert any relation between truth and present success.

Let me move to the second reply available to the teleosemanticist.

Multiple causes The second difference between cases of the dormitive virtue-style accounts and teleosemantics needs a bit more explaining. Shea thinks the teleosemantics suffers from the circularity problem because of its functionalist dimension. If one defines property R by appealing to the fact that it causes (or it is disposed to cause) B, then one cannot *explain* the occurrence of B by appealing to R. If, for instance, one defines *being in pain* as the property that causes certain avoidance behavior, one cannot explain someone's avoidance behavior by saying that she is in pain. Shea adds that a standard move on behalf of functionalism is to specify the property using multiple dispositions. For example, if one says that pain is not just the disposition to certain avoidance behavior, but it is also caused by bodily damage, leads to anxiety and so on, then instances of the disposition can be picked out without observing the effect one is trying to explain. Accordingly, if one defines pain by using a complex network of dispositions, the claim that someone is in pain can provide a more substantive explanation of a particular manifestation, like avoidance behavior.

Shea claims this is precisely what his theory does. By adding an input condition, he argues that his theory solves the circularity problem because a semantic property is now ascribed by appealing to multiple features. Since carrying content C is defined by a complex property that includes many different conditions (including its carrying information), by mentioning this complex property one can provide an explanation of one of the manifestations of one disposition.

But this kind of response seems also to be available to teleosemantics. According to teleological theories, content depends not only on producing successful behaviors; there are many other conditions involved. There must be two systems, with certain etiological functions,

35 This backward-looking feature of teleosemantics has also originated some of the most serious objections, like Swamp-cases: any trait with the wrong history cannot have functions and, hence, its states cannot be representations (see 3.3.4). This is a difficulty that, for instance, dispositional or counterfactual theories can easily deal with (Abrams, 2005; Fodor, 1990)

they must have co-evolved as cooperating systems, and so on. In that respect, there is a significant difference between success semantics and the elaborated conditions that teleosemantics requires in order to carry certain content. Hence, when a semantic property is attributed to a certain state, one is implicitly assuming that many other facts and conditions (besides usually leading to successful behavior) obtain. That confers a very important explanatory value to content attributions.

Productivity Finally, I think Godfrey-Smith's and Shea's objection clearly fails when we focus on a teleosemantic theory that be able to account for productive representations. The circularity argument simply collapses when we think about productive representational systems. According to teleosemantics, asserting that bee dance nr. 873 is true is not merely to subsume this bee dance under a pattern of content-constituting situations, because probably no bee dance had previously had this particular content. The fact that this particular bee produces a true representation of that particular state is a notorious achievement. It means that certain mechanisms that lead to true representations in the past also have led to a true representation in that occasion. So when a teleosemanticist claims that a representation caused a successful behavior it says much more than 'the current case falls into the same pattern as the past cases that were content-constituting' (Shea, 2007, p. 12). As I suggested earlier, I think Godfrey-Smith and Shea have probably been misled by the simplified version of teleosemantics (along the lines of First Teleosemantics), and have not considered more complex versions that can accommodate productive systems, like Third Teleosemantics.

Therefore, the fact that teleosemantics can account for new contentful representations clearly illustrates the fact that attributing a true representation cannot be the same as saying that a certain state of affairs that used to occur in the past also occurs now. In many cases, teleosemantics attributes (true or false) contentful representations even if no representation with that particular content has ever existed.

Finally, if we put these three aspects together (the backward-looking dimension, the appeal to multiple causes and the version of teleosemantics that accounts for productivity), we will easily see why the semantic properties attributed by teleosemantics are really explanatory. By ascribing a semantic property, we are assuming that a whole range of facts obtain (related to systems, etiological functions, coevolution...) and, crucially, among them the success of the actual behavior is not included. These different features show that teleosemantics is perfectly compatible with content ascriptions being really explanatory.[36]

36 One might agree that this is a good reply for explanations of current successful actions, but object that this reply fails if we focus on certain explanations of past success. In particular, she might reply that we want semantic properties to explain not only why *current* representations usually prompt certain behaviors; we also want to be able to say that *in the evolutionary past* the fact that certain organisms had beliefs partially explained their behavior. But (the objection runs) according to teleosemantics, the semantic properties of *past* instances of a given representation are determined by the successful output of *past* representations, so at least in this respect, circularity threatens. Fortunately, the answer I provided for current situations can also be employed in these other cases. At any time t, the fact that the system produces a representation of a certain state of affairs is determined by certain facts that happened before t. So for any time t, when we explain the organism's success by appealing to its representational capacities, we are not considering the probability of succeeding at t, but something that occurred before t.

Summing up, Third teleosemantics does not render explanations of successful behavior in terms of having certain contents circular. Third Teleosemantics certainly entails that the fact that an organism has a belief with a certain content is explained by appealing to certain facts related to successful behavior. But the relevant behavior is that of one's *ancestors*, and the fact that a given state has a certain content is not based on whether it will led to successful behavior in the current situation. Moreover, an ascription of a mental state with a certain content assumes a whole range of issues concerning the existence of the right systems, its functions, etc. Finally, if we consider the standard version of teleosemantics that is able to account for the productivity of certain systems, its explanatory import is even more obvious. Therefore, if Shea's objection is interpreted as involving a Circularity of Content, the conditional in premise (2) of his argument turns out to be false: having a belief can be explained by appealing to (past) successful behavior and nevertheless it can satisfactorily explain (current) success. As a consequence, premise (3) ('according to Teleosemantics, true belief cannot explain success') turns out to be false, and the argument does not go through.[37]

circularity of true content Let us now consider the second way of interpreting the objection: the Circularity of True Content.[38] Prima facie, if there is no circularity in appealing to beliefs in order to explain successful behavior, there is no reason why *true* beliefs should fail to be explanatory either. So, if our previous response to the circularity of content was on the right track, there is no reason why the attribution of true beliefs should be problematic.

In that respect, notice that the *truth* of a representation is not explained by successful behavior (past or present); not even in teleosemantics. Teleosemanticists usually adopt a correspondentist theory of truth (as Shea does), according to which (roughly) a representation is true iff the represented state of affairs obtains. So, strictly speaking, premise (1) of Shea's argument ('having a true belief is explained by successful behavior') is only right to the extent that having a belief is partially explained by past successful behavior and we just saw in the last section that this is not objectionable. Since the fact that a representation is true is not explained by actual or past behavior, it should be clear that Circularity True Content does not threaten Third teleosemantics.

Indeed, it could be argued that even if there were some circularity in the definition of belief (i.e. if Circularity Content were right) the claim that a belief is true could provide a substantive explanation according to Teleosemantics. After all, the claim that John has a true belief that p is just the claim that John believes p and p obtains. Its being a true representation is explained by the fact that (1) the organism has a

37 Let me mention that an aspect that might have confused Shea is that in normal conditions an attribution of a semantic property *entails* an attribution of certain dispositional properties. For instance, entertaining a certain belief entails that one is disposed to act successfully if certain normal conditions obtain. But this is a usual *consequence* of property ascriptions: in general, an attribution of any property P usually implies an attribution of a dispositional property D, and nevertheless, the fact does not imply that one can not satisfactorily explain D by mentioning P. The fact that I have a nose with a certain shape entails that in normal conditions it has the disposition to support glasses, but I can surely explain my nose's capacity to support glasses by appealing to its shape.

38 This interpretation is suggested by quotes like the following: 'The issue is rather whether *true representation* can explain success (...) That is, the question is whether statements of the following form can explain successful behavior of a system S: 'p and S represents that p''. (Shea, 2007, p. 415. Emphasis in the original)

representation with a certain content in virtue of the selective story of its ancestors and (2) the content is satisfied. So, surely, the fact that an organism *currently* has a true belief is not explained by the fact this organism *currently* behaves successfully. Even if the ascription of content involves some circularity, its truth would have a significant explanatory value.

CONCLUSION In conclusion, I think Godfrey-Smith's and Shea's argument against standard version of teleosemantics is flawed. THIRD TELEOSEMANTICS does not render explanations of successful behavior in terms of true content circular. Furthermore, I have argued that INFOTELSEMANTICS fares no better (and probably worse) than teleosemantics. So I think Shea's and Godfrey-Smith circularity challenge to standard versions of teleosemantics has been met.

Let us consider now whether THIRD TELEOSEMANTICS is right in assuming a cooperation between sender and receivers.

3.3.3 *The Cooperation requirement*

Condition 2 in THIRD SENDER-RECEIVER claims that a Normal condition for the performance of P's and C's function is the presence and proper functioning of the other mechanism. That means that the fact that the producer P has performed its function helps to explain why the consumer C historically complied with its function, and the fact that C performed its function helps to explain why P historically complied with its function. That is the relevant sense in which P and C must be cooperating devices.

Why should we think producer and consumer systems have been cooperating? The intuitive idea behind this claim (which is also supported by abstract models of signaling systems such as Lewis, 1969; Skyrms, 1996) is that, on the one hand, senders acquire the capacity of producing signals only if they are benefited from the receiver's activity; otherwise, they would stop producing signs (Millikan, 2004, 2005). If the sender did not benefit from the action of the consumer system, it seems it would not evolve a mechanism for informing the receiver about the presence of some state of affairs. Similarly, receivers must benefit from the senders performing their functions; otherwiese, they would ignore the sign (Shettleworth, 2010, p. 513). That shows that a Normal condition for the proper performance of each system (producer and consumer) is the presence and proper functioning of the other. In other words, sender and receiver must have coevolved as cooperating systems.

The idea that sender and receivers should cooperate is entrenched in scientific reasoning as well:

> If there is, on average, no information of benefit to the receiver of a signal, then receiver should evolve to ignore that signal. If receivers ignore the signal, then signaling no longer has any benefit to the signaler, and the whole communication system should disappear. (Searcy and Nowicki, 2005)

> An analysis that allows the signaler's behavior to evolve but does not permit any evolution in receiver's response does not make sense (...). In fact, receivers should evolve

> responses to signals only when it is advantageous to do so. And if it does not benefit receivers to respond in a particular way to a specific acoustic feature then selection will favor receivers that attend to some other cue. (Seyfarth et al., 2010)

A second reason for thinking that cooperation should be a requirement derives from the way content is determined according to teleosemantics (specified in THIRD CONTENT). The theory claims that the content of a sign produced by a sender is determined by the historical needs of the consumer. So, if P is a sender and produces a state R that represents S, then by definition there must be some consumer that has performed its function usually enough thanks to the presence of S. Consequently, if P is a sender that produces meaningful signs, there must be a consumer that usually enough benefits from perceiving the sign. The way content is determined according to teleosemantics seems to entail that representations originate between systems that have at least partial common interest.

In that respect, it is important to stress that the requirement of cooperation does not demand complete common interest; it suffices if the sender and the receiver have partial common interest. Each one must somehow benefit from the activity of the other, and that benefit must partially explain the selection of the mechanism.[39] As condition 2 in THIRD SENDER-RECEIVER claims, the Normal condition for the proper performance of each system is the presence and proper functioning of the other. The contribution of each, however, might be suboptimal (some models of partial interest can be found in Skyrms, 1996, 2010).

Despite the intuitive plausibility of this claim, some people have recently argued that condition 2 should be dropped from the theory. Their criticism is motivated by some cases that apparently illustrate the existence of signaling without cooperation. Let us consider this objection in some detail.

3.3.3.1 *Uncooperative systems*

The cooperation paradigm has been attacked at the same time by many biologists and philosophers. From the scientific domain, the idea that signaling must assume a certain degree of cooperation among participants has been seriously challenged from a general perspective on evolution (Dawkins and Krebs, 1976) as well as from a different ethological approach (e.g. Rendall et al., 2009). Here I will leave aside the general problem of selfishness and manipulation (which, I think, has already been sensibly replied from an evolutionary perspective by Godfrey-Smith, 1996,2009 and from an ethological point of view by Sefarth et al., 2010) and I will focus on a set of counterexamples that philosophers have raised against the cooperation requirement included in the teleosemantics framework.

39 It is important not be misled here by the different uses of the expression 'common interest'. For instance, Maynard Smith and Harper (2003, p. 27) define cases of 'common interest' as involving two organisms that 'place the possible outcomes of the interaction in the same rank order of preference'. This is a stronger notion from the one I am using here (and the one that is required for teleosemantics). In the sense intended here, there can be common interest between two organisms even if there is partial competition or even if signaling involves some partial cost that could be avoided by the organisms involved. This is why, prima facie, phenomena like the 'handicap principle' (Zahavi, 1975; Zahavi and Zahavi, 1997) do not threaten teleosemantics.

In this context, the objection was originally raised by Sterelny (1995) and developed by Stegmann (2009) and it is based on the phenomenon of mimicking. Mimicry (or mimetism) is the similarity of one species to another which benefits one or (less frequently) both. Usually, this similarity is in appearance, scent, behavior or sound. There are different kinds of mimicry; sometimes separate unpalatable or dangerous species evolve similar appearances in order to reinforce the warning signals that predators can learn (*Müllerian mimicry*). In other cases, organisms mimic certain species in order to look more dangerous (*Batesian mimicry*). For instance, the Ash Borer (*Podosesia syringae*) is a Batesian mimic of the Common wasp, because it has copied the wasp's black-and-yellow strips, but it is unable to sting. Similarly, the Viceroy butterfly (*Limenitis archippus*) mimics and it is mimicked by the Monarch butterfly (*Danaus plexippus*); since both are to some extent unpalatable butterflies, they both benefit from the fact that predators confuse them (but see Ruxton et al. 2004, ch. 9).

More precisely, the objection to teleosemantics is based on what ethologists call 'agressive mimicry' (Marshall and Hill, 2009, Eberhard, 1977). In aggressive mimicry, a predator or parasite imitates the signal of another species in order to exploit the recipient of the signal. A species of Australian katydid (*Chlorobalius leucoviridis*), for example, imitates the mating sound of female cicadas in order to attract male cicadas and devour them (Marshall and Hill, 2009). Similarly, the bolas spider (*Mastophora* species) attracts male moths by imitating the sex pheromones of female moths (Eberhard, 1977). All these cases seem to involve signaling without cooperation.

Now, in order to develop the objection in some detail, let me focus on the case of aggressive mimicry in fireflies, which is Sterelny's original example, and the one Stegmann (2005, 2009) appeals to in order to spell out his criticism:

> Among the fireflies are some species that prey on other fireflies. Females of the species *Photuris versicolor*, for example, prey on the males of several *Photinus* species. Predation involves the deceptive use of mating signals (Lloyd, 1975). The aptly named 'femmes fatales' lure the males by sending the sort of mating signals that the males' conspecific females would send. So, for instance, if a predator perceives the flashes of a male *P. macdermotti,* and if she's hungry, then she will emit the sort of flash that a female *P. macdermotti* would emit if she were willing to mate. (...) From an ethological point of view, the predator's female-*macdermotti*- type flash carries the information that there is a female *P. macdermotti* willing to mate.(...) But the co-occurrence of a female *macdermotti* type flash with the presence of a hungry predator is clearly not the normal condition for the male's consuming device to achieve its function. (Stegmann, 2009, p. 868)

Let us try to describe more carefully the case having in mind the Sender-Receiver framework we have been working with (for simplicity, let us call members of the Photinus species 'F-females' and 'F-males' and members of the Photuris versicolor species 'Predator'). First, if we set aside for a moment the parasiting behavior and focus on the usual behavior of the F-species, the schema described in Third Teleosemantics happily applies. F-females (producer) Normally send a signal (light)

to F-males (consumer). Since the behavior of the F-males (i.e. mating) was historically successful only in those occasions where there was a F-female ready to mate and because there was such a female, then Third Teleosemantics predicts that the light emitted means something like *F-female willing to mate.* So far so good.

The problem arises when we focus on the parasiting behavior. The light emitted by the 'femmes fatales' of the Predator species (*Photuris versicolor*) seem to be a representation. Indeed, it intuitively means the same as the light emitted by the parasited bug, namely something like *F-female willing to mate.* This is the reason F-males are attracted to Predators, which do not hesitate in devouring them. In fact, it seems that only if we assume that the light emitted by Predator has the same content as the light emitted by F-females (the parasited bug) can we explain the behavior of F-males. Hence, the following claim seems to be true (and widely assumed by ethologists):

Mimicry: The light emitted by Predator is a representation, which means something like *F-female willing to mate.*[40]

The key problem Sterelny and others point out is that it seems Third Teleosemantics cannot accommodate Mimicry. First of all, notice that the receivers of the light emitted by Predator are the F-males, so in this case the sender and the receiver are instantiated in two organisms that constitute predator and prey. Since we can reasonably assume that in this case predator and prey have no common interest,[41] it seems Mimicry entails that a state can be a representation even if the sender and the receiver are not cooperating devices. That clashes with condition 2 of Second Sender Receiver.

Furthermore, notice that the content of the light signal emitted by Predators does not seem to be the state that the consumer has historically needed in order to perform its function in a Normal way when the signal was present. The content of the light seems to be *F-female willing to mate*, but if we look at the past cases in which Predator signals where produced, there were no F-females willing to mate, but only hungry Predators. Surely, nothing like the presence of a F-female willing to mate obtained in the historical circumstances that explain the selection of the producer system of Predator and the consumer system of F-males.

Consequently, the following claim seems to be true:

Incompatibility: Third Teleosemantics is incompatible with Mimicry.

We have, then, three plausible thesis that are mutually inconsistent: Third Teleosemantics, Mimicry and Incompatibility. If one accepts

40 Mimicry is defined in terms of the *Photuris versicolor* and *Photinus* species, but notice that the problem pointed out here concerns (at least) any case of aggressive mimicry. This example is supposed to highlight a broad and significant set of cases that teleosemantics cannot deal with.

41 It has been argued that, in some cases, predator and prey may have some common interest. For instance, according to the Perception Advertisement Hypothesis, some organisms inform their predators that they have been perceived, so that hunting per surprise becomes futile (Radner, 1999, p. 129-130). Gazelles, for example, perform a set of controlled jumps (called 'stotting') so as to communicate to the predator that it has been detected (Sterelny and Griffiths, 1999) or that it is a healthy exemplar (Maynard-Smith and Harper, 2003, p. 61; Ruxton et al. 2004, ch. 6). Apparently, this sort of signs benefit both predator and prey; the former does not attempt an attack that will probably fail and the latter avoids a possible threat (Millikan, 2004; Ruxton et al. 2004, ch. 6).
Even if these examples exist, aggressive mimicry seems to be a different sort of case. It is extremely plausible that the light emitted by Predator in order to lure F-males only benefits the former.

Third Teleosemantics and Incompatibility, then Mimicry should be rejected. If, on the contrary, one holds Mimicry and Incompatibility, then Third Teleosemantics should be given up. Finally, if one wants to maintain Third Teleosemantics and Mimicry, Incompatibility must be abandoned. At least one of them should be given up. Let us consider these options in some detail.

A first option is to hold that this counterexample suggests that the whole framework set up in Third Teleosemantics must be entirely rejected. That is probably an extreme position to take, since Third Teleosemantics seems to yield the right results in a wide range of cases and has independent support.

A more refined and popular version of the this first option consists in modifying Third Teleosemantics in order to make it compatible with Mimicry . For instance, one could argue that Third Teleosemantics specifies a set of sufficient but not necessary conditions for representational systems to arise (along the lines of Sterelny, 1995; Sterelny and Griffiths, 1999). Defenders of this proposal seem to be committed to a 'splitting account' of the phenomenon of representation, according to which different sorts of representations require different analysis. Another strategy is to alter the sender-receiver structure described in Third Teleosemantics. Stegmann (2009) and Cao (2012), for instance, suggest that content is only determined by the consumer system. According to them, coevolution and cooperation is not required; what a state represents depends only on the state of affairs that a consumer systems needs. Let me point out, however, that if the arguments for the cooperation requirement suggested earlier are sound, it is not clear these proposals are coherent with the main insights of teleosemantics (for instance, how content is determined) and it seems they will probably clash with the intuitive claims presented in 3.3.3 and defended by many ethologists.

I will defend a different option. I think that Mimicry and Incompatibilty can be rejected. Of course, in order for standard teleosemantics to overcome the difficulty, it suffices if one of them is abandoned. However, I think it is important to show that there are many options available to the teleosemanticist. I will argue that cases of aggressive mimicry can be perfectly accommodated within Third Teleosemantics, either by denying that Predators really emit signals, or by showing that the fact that they send signals is compatible with the theory. I will not try to argue which of these options is more plausible. The task of the remainder of the paper is to show that parasitic behaviors like the one depicted earlier are not in tension with Third Teleosemantics.

3.3.3.2 *Accounting for Uncooperative Mechanisms*

rejecting mimicry Let me start by considering the more straightforward way of solving the puzzle. The first strategy is to reject Mimicry and maintain that, strictly speaking, Predator does not produce representations, but meaningless states. This option assumes that the light emitted by Predators are not really signals, even if they look exactly like the signals emitted by F-females. More generally, the first suggestion is that in cases of aggressive mimicry in which a sign is copied, no real signal is produced by the mimicker.

There is of course an obvious problem with this proposal. The teleosemanticist could be accused of offering an ad hoc solution to a serious objection. Is there any reason (besides rescuing teleosemantics)

for thinking that the Predator's flashings are not really representations? After all (the objection runs) they resemble very much the original signals and have the same effects, i.e. attracting F-males. Furthermore, it seems that ethologists usually explain the behavior of F-males by assuming that the light emitted by Predator are representations that mean something like *there is an F-female willing to mate*. If ethologists explain this parasiting behavior by appealing to the meaning of the signal, it seems we have a prima facie reason for thinking that it is indeed a signal.

In response, there are at least two important considerations besides teleosemantics for rejecting Mimicry. The first one concerns the explanation of behavior and the second one has to do with general explanations of mimicry and cryptic strategies.

First of all, I think that a careful look at this sort of examples shows that an explanation of this case does not require assuming that the light emitted by Predator is really a signal. One can perfectly accommodate this situation by merely assuming that F-males *wrongly think* (or, to use a less cognitively loaded term, *represent*) that the Predator's light is a signal. And, of course, there is a good explanation for this confusion, based on the strong resemblance of the light emitted by Predators and F-females. In other words, we can fully explain the phenomena by saying that F-males are simply wrong; the light produced by Predator is not a signal and does not mean anything, but there is a simple explanation of why F-males are misled. The key element in the explanation of the behavior of F-males is the fact that they act as if that the light emitted by Predators was a signal (see El-Hani et al., 2010, p. 11). The additional claim that this state is indeed a real signal sheds no additional light onto this explanation.

Considering other cases of aggressive mimicking might help to clarify this point. Think about the astonishing example of the Blister Beetles (*Meloe franciscanus*). Just after hatching, larvae of the blister beetle climb to the top of stems where they form an aggregation that resembles a bee. These aggregations attract (through visual and chemical cues) male bees, which try to copulate with them. During the pseudo-copulation, larvae attach to the male bee and are eventually transported to the bee's colony that they will parasitize (Hafernik and Saul-Gershenz, 2000). This is usually classified as an example of aggressive mimicry (Ruxton et al, 2004, ch. 6). However, in this case the hypothesis that male bees are misled because an aggregation of blister beetles' larvae really constitute a female bee is preposterous. In general, we do not expect mimicking and mimicked entities to be of the same kind. Male bees are misled into thinking that there is a female bee ready to mate because an aggregation of larvae look and smell like them, but they are simply wrong. Similarly, F-males are mislead into thinking that the Predator's light is a signal, but they are wrong. Consequently, Predators do not emit real signals, but only flashings that resemble signals. As a result, Mimicry turns out to be false.

A second reason for rejecting Mimicry is that taking this perspective has interesting advantages from a scientific point of view. There are many strategies organisms employ in order to confuse others. For instance, in the phenomenon known as 'masquerade', organisms tend to resemble innanimated things in order to be avoided by predators. In contrast to the strategy of background matching (that is, standard cases of camouflaging), in masquerade the organism is usually detected

but confused for another thing. A remarkable example includes the sea dragon (*Phyllopteryx eques*), an Australian see-horse with numerous outgrowths that resembles a see weed (Ruxton et al. 2004, p. 23). Likewise, some Amazonian fish species avoid predators by resembling dead leafs (Sazima et al, 2006). A similar phenomenon is the so called 'disruptive colloration', in which the organism's coloration tends to obscure the true form of the animal and conceal certain parts. For instance, it has been suggested that the white spots on the morph of the isopod *Idotea baltica* serve to obscure its real form rather than to match spots in the background (Merilaita, 1998). Another strategy is deflection which works by increasing the predator's probability of striking at a highly defended or expendable part. Some lizards, for example, have brightly colored tails, which contrast with the cryptic coloration of the rest of the body. This conspicuous color increases the likelihood of an attack being directed at the tail, which lizards can shed and regrow (Ruxton et al. 2004, p. 183). We could also add to this list cases of Batesian and Müllerian mimicry, explained above.

Now, intuitively, there is something important that all these strategies have in common: their function is to lead predators to *misidentify* the prey. That is, what explains the evolution of all these strategies is that often enough they manage to produce false representations in predators. Predators think (or represent) there being a see weed, or there being a leaf, or there being a blurry entity with unclear contours. This is the central function that explains why all these different forms of camouflaging and mimicking have evolved. Classifying them together has obvious advantages from a scientific point of view. Despite the significant differences among these strategies, some models and generalizations are applicable to all of them, so highlighting this common background has fruitful consequences for some research programs (Ruxton et al. 2004).

Aggressive mimicry is usually understood within the same paradigm. For instance, many of the models and theories that are useful for explaining cases of Batesian mimicry or masquerade can also be employed in explaining aggressive mimicry (Maynard-Smith and Harper, 2003). Therefore, from a scientific point of view, it makes a lot of sense to focus on the fact that the function of all these strategies is to mislead predators. This claim lends support to the idea that the central explanatory notion is that of misidentification.

Consequently, an interesting scientific perspective classifies most cases of camouflaging and mimicry by appealing to the fact that they lead other organisms to misrepresent. What unifies all these strategies is that they cause misidentifications, not that they are signals. I also argued that the claim that the light emitted by Predators is a signal is not doing any explanatory work and that, in general, we do not expect mimicking and mimicked entities to literally be members of the same kind. As a result, I think there are good reasons for rejecting Mimicry.

REJECTING INCOMPATIBILITY I just argued that one option is to reject Mimicry and maintain that, strictly speaking, the mimicking system does not produce representations, but meaningless states. A second strategy I would like to discuss is whether one can endorse Thid Teleosemantics and Mimicry and reject Incompatibility. The goal is to argue that one can coherently hold that the Predator's light is a representation that means something like *there is an F-female willing to*

mate and, at the same time, that cooperation between producer and consumer is a requirement for a state to qualify as a representation.

In what follows, I would like to show that, if one assumes Thid Teleosemantics and Mimicry , there are two different ways in which teleosemantics can accommodate cases of aggressive mimicry: copying signals and copying mechanisms.

Copying signals Let us grant Mimicry for the sake of the argument, that is, let us assume that the signals emitted by Predator are indeed representations. If one grants that much, it is should be obvious that the signals emitted by Predator have the same content as the light emitted by the mimicked females (F-females). Of course, this is not a mere coincidence; the content of the light signal of Predator seems to completely depend on the content of the light emitted by F-females. If the representational content of the signal produced by F-females were different (e.g. *there is food nearby*), we would conclude that the content of the signal of Predator would change accordingly. This is a point that needs explaining.

Secondly, notice that not only the content of the representation, but the non-intentional properties of the signal itself (light intensity, brightness,..) entirely depend on the features of the parasited representational system. The representational system of the mimicking system must resemble as much as possible the representational system of the mimicked organism. There is a strong tendency to copy any feature of the parasited sign. If the intensity of the light emitted by F-females were to change, there would probably be a strong tendency for Predator to change the intensity of their signals accordingly. In fact, not only the physical properties of the signal are imitated, but also some of its functions (Stegmann, 2009, p. 871-2). All flashes have the function to attract F-males. Consequently, the similarities between the parasiting and parasited systems seem to be deep and well grounded.

Indeed, this relation of dependence is not accidental, because the properties of the parasiting representational systems are (historically) caused by the properties of the parasited system. Proof of that causal relation is that if the parasited representational system were to change in relevant aspects, the parasiting system would also change accordingly. The fact that there is a counterfactual dependence relation between one and other suggests that there might be a causal relation between them (Sober, 1984). This causal relation to a great extent explains the commonalities between the mimicking and the mimicked system. The mimicked and the mimicking signaling systems tend to have the same properties because the mimicking system is the result of a causal process of copy. This is a significant result.

On the other hand, remember that a common way of individuating kinds in biology is precisely by appealing to this kind of causal process of copy. In particular, according to a very popular theory defended by Boyd (1999a, 1999b), Griffiths (1999), Wilson (1999) and Millikan (2000) among others, biological kinds are groups of entities that share stable similarities due to an underlying causal process. In a nutshell, the idea is that two entities belong to the same kind if they tend to resemble each other in important respects due to an underlying process of copy. i.e. if they belong to the same Reproductivley Established Family (REF, see 3.2.4).

These ideas naturally lead to the first proposal: the flashings emitted by Predators are signals because they belong to the same REF as the light emitted by the parasited bugs. That is, the Predators' flashings are signals in virtue of the fact that they tend to resemble the signals emitted by F-females due to an underlying robust causal mechanism. There is a strong evolutionary tendency for the Predator signal to reproduce any properties of the signal of F-females, and this process of copy is enough for justiyfing the claim that the signal emitted by Predator and the signal emitted by F-females belong to the same type of signals. Many properties like the intensity of the light, its brightness, its frequency and so on are copied through a robust causal process. So, one could reasonably argue that, at some point, this process of reproduction justifies the claim that the Predator's light is indeed a signal which also copies the meaning from the original. Both the signal of the mimicking and the signal of the mimicked organism are of the same kind in virtue of that process of copy.

The proposal, then, is that a popular perspective on the nature of biological kinds has the consequence that the Predator's flashings and the F-female's flashings are both signals of the same kind. Notice that this approach can explain why the content of the Predator's signal is exactly the same as the content of the F-female's signal and, moreover this explanation seems to be compatible with Third Teleosemantics. We saw that Third Teleosemantics can easily explain why the signals of F-females directed at F-males are representations. In order to account for the Predator's flashings being signals, we just need to accept that they ride piggyback on the signals of F-females.

Copying mechanisms Before presenting an possible objection to this approach, let me outline a different way in which Mimicry and Third Teleosemantics can be said to be compatible. There is a second strategy for showing that Third Teleosemantics is fully compatible with Mimicry. One could try to argue that it is not just that the mimicking and the mimicked flashes are both signals of the same type. A more ambitious hypothesis is that, in order for Third Teleosemantics to account for cases of parasitism, we must simply realize that parasitic representational *systems* belong the the same biological kind as their parasited *systems*. In other words, both the producer system of F-females and the producer system of Predator belong to the same biological kind. The proposal, then, is that the parasiting mechanism system is a mere copy of the parasited one. This surprising idea is supported by two claims: on the one hand, the thesis that two entities belong to the same kind if they tend to have important properties in common in virtue of some robust causal process of copy. On the other, the observation that this strong causal process of copy is taking place in the case of the signaling system of F-females and Predators. Since we might think this is a robust and non-accidental link that has been active during the evolution of the whole representational system, the producer systems of F-females and Predator could satisfy the criteria for qualifying as members of the same biological kind.

Now, if we accept that the producer system of Predators and the producer system of F-females belong the same biological kind, then Mimicry can be perfectly accommodated within Third Teleosemantics. What explains the perplexity is that we were previously misapplying the sender-receiver framework to the case of fireflies. Given that

producer and consumer systems must constitute biological kinds (after all, they must be selected for) and given that the light-emitting mechanism of F-female and Predators belong to the same kind, in order to apply the sender-receiver framework properly we should be assessing whether the light emitting mechanism of F-females *and* Predators has some common interest with the consumer system of F-males. And, once the question is cashed out in these terms, the answer seems to be clearly affirmative.

I admit that, at first glance, this proposal might look unpromising. In particular, it might seem that if the producer system is constituted by the light emitting mechanism of F-females and Preadors, then this producer probably does not satisfy the conditions set up in Third Teleosemantics. However, I think that, if one looks cerfully at the teleosemantic framework, this sensible concern turns out to be ungrounded.

Consider the following question: Is the function of P (which includes the signaling systems of Predators and F-females) to produce a state R when another state obtains (in particular, the state *F-female ready to mate*)? Yes, it is. The light producer P (which includes the mimicking system of Predator) has the function of producing a flash when there is a F-female ready to mate. But justifying this claim requires some elaboration.

On the etiological understanding of functions, functions are selected effects. That is, the function of a trait is the effect that explains why past tokens of this trait were selected for. Now, the explanation of the selection of producers P relies on the fact that usually enough light emitted by F-females corresponded to an F-female ready to mate. Crucially, only the light emitted by F-females (and not the flashed emitted by Predator) helps to explain the existence of the representation system; the producer system in Predator rides piggyback on the success of the system in F-females. In other words, the producer P in firefly signaling exists *despite the fact* that this kind includes Predators, which reduces the overall reliability of the whole representational system. What causally explains the selection of mechanisms P is the presence of F-females ready to mate. Signaling systems in Predator just take profit and copy the system in F-females; and, since they do not positively contribute to the selection of the whole mechanism, the activity of Predator does not alter the function of the representational systems. Consequently, even if the the producer system of Predator and the producer system of F-females belong to the same REF, its function (and content) is exclusively determined by the effects of F-females and F-males.

Now, given that the function of P and C is not altered by the presence of some P that do not contribute to the overall fitness, it seems that the cooperation requirement of third teleosemantics (condition 2) is also satisfied. The Normal condition for the proper performance of each system is the presence of proper functioning of the other. There is indeed partial common interest between producer P and consumer C.

Therefore, this is a different way of showing that both Mimicry and Third Teleosemantics are compatible. So the following two claims can be true at the same time: (1) members of Predator produce a representation that is consumed by F-males and (2) cooperation is required for a state to qualify as a representation. The key suggestion that dissolves the perplexity is that the sender-receiver model (and the cooperation requirement) applies at the level of kinds of mechanisms,

and at this level the mechanism of F-females and Predators may belong to the same kind.[42]

Notice that assuming that the Predator's flashings are signals in virtue of being copied and assuming that the Predator's flashing are signals in virtue of being produced by a mechanism that is a copy of the producer of F-females lead to two very different solutions to the problem. According to the first proposal (*copying signals*), teleosemantics applies at the level of F-females and F-males and parasite's flashings simply ride piggyback on these signals. In contrast, this second solution (*copying mechanisms*), assumes a different way of typing systems; the producer systems in Parasites and F-females form a single kind, so that common interest between producers and consumer is justified.

As a final remark, it is worth stressing that classifying the producer of Predators and the producer of F-females as belonging to the same biological kind would have significant consequences in some areas of biology. This is an important issue that deserves to be seriously taken into account before this last proposal is utterly adopted or rejected. Nonetheless, my goal was merely to show that there are many options available to the teleosemanticist in order to accommodate cases of mimicry. Much more should be done in order to show that cases of aggressive mimicry pose a significant problem for teleosemantics

Summing up, I have considered three ways in which teleosemantics can account for cases of aggressive mimicry. The first one is to reject Mimicry and hold that the light emitted by predators is not really a signal. The second strategy is to accept Third Teleosemantics and Mimicry, but deny Incompatibility. I have shown there are two ways of doing that, either by assuming that flashings are of the same type or assuming that mechanisms are of the same type. To complete this discussion, let me now turn to Stegmann's objection.

3.3.3.3 *Stegmann's reply*

In his recent paper on that issue, Stegmann (2009) seems to shortly consider the reply based on the rejection of Incompatibility. In particular, he writes:

> Might the predator's flashes have content because they inherit it from the cooperative flashes they mimic? The notion of copy of 'reproduction' plays an significant role in Millikan's (1984) account. The predators' flashes, however, do not qualify as 'reproductions' in her technical sense. 'Reproductions' share properties with the model due to the fact that the model is directly causally responsible for the reproductions' properties (Millikan, 1984, p.20). Imitiations like the parrot's 'hello' are reproductions in this sense. But there is no such direct causal link from cooperative to mimicking flashes. Nor do the predators's flashes form a 'higher-order reproductively established family' together with the females'

42 Of course, there is a sense in which the function of the signaling system in Predator is to *prey* on F-males, while the function of the system in F-females is to *mate* with F-males, but there are many ways this fact can be accommodated. First, the same trait belongs to many REF and, accordingly, can have many functions at the same time. Secondly, it is possible to accept that the function of the particular mechanism that produces light signals in Predator is to do one thing, and at the same time hold that this mechanism is included within a larger system (perhaps a 'prey detecting system', which includes other subsystems) that has a different function. Thus, this solution to the problem of uncooperative systems is compatible with Predator having different goals from F-females.

> flashes. For this would require that either all flashes are produced by the same device or, if produced by distinct devices, the devices are reproductions of one another (Millikan, 1984, 24-5). Neither is the case. (Stegmann, 2009, p. 869, emphasis in the original)

I think this reply fails to provide a convincing argument. It seems hard to deny that the flashes of Predators are reproductions of the flashes of F-females. They have the same intentional and non-intentional properties (intensity, frequency..), they share certain functions and they have these properties in common for a reason. Furthermore, the systems of Predator that emit these flashes have been designed by evolution in order to match the signals emitted by F-females. That much seems to be pretty uncontroversial. The question now is the following: Why are these similarities and causal relations between signals and producer systems insufficient for establishing the relevant 'reproductive relation' between signals or even between systems? Stegmann seems to assume an unwarranted too narrow understanding of 'reproduction'.

There is a different reason for thinking Stegmann's reply is mistaken. The problem of uncooperative systems that we are discussing assumes that the content of the mimicking signals is the same as the content of mimicked signals (this is presupposed in MIMICRY). But, the only way of explaining this fact is by assuming that both signals belong to the same type. How else could we explain that the content of the signal of Predators is F-female ready to mate?

Interestingly enough, Stegmann (2009, p. 871) himself claims that the mimicked and the mimicking signal belong to the same type (what seems to be in tension with the previous objection to teleosemantics):

> The first condition [of Stegmann's account] endows female-macdermotti-type flashes with representational content irrespective of whether they were generated by females or predators.

Since Stegmann assumes that the signs emitted by mimicker and mimicking systems belong to the same type, we might wonder why teleosemantics cannot accept that. In fact, as I said, I think this is an assumption that any plausible account of aggressive mimicking should make. But once we accept that the signals of Predator and F-female belong to the same kind, then we are naturally led to one of the two solutions I gave.

Therefore, I think Stegmann's objection against the kind of proposal I just offered is flawed.[43]

CONCLUDING REMARKS In this section I have argued that examples of parasitic representational systems do not constitute a counterexample to the idea that representational systems require certain amount of cooperation between sender and receivers. I have argued that there are three different ways of accommodating cases of aggressive mimicry within the theory. If the arguments I presented are on the right track, teleosemantics can explain cases of aggressive mimicry with the same

43 Indeed, I think Stegmann's own proposal (developed in Stegmann, 2009) accounts for cases of aggressive mimicking, not because he modifies THIRD TELEOSEMANTICS, but because he assumes that all signs belong to the same kind and hence have the same content. However, I have just argued that once we accept the latter, THIRD TELEOSEMANTICS can also offer a satisfactory reply.

framework that it uses in the rest of representational systems. That shows that a set of cases that most people thought were problematic for teleosemantics are fully compatible with standard version of the theory, which assumes that sender and receiver must be cooperating systems.

3.3.4 *Swampman*

The problem of Swampman has accompanied Teleosemantics since it originally appeared in the philosophical scene around 1984. So I think it points at a significant consequence of the theory that any serious approach must address.

Contrary to common wisdom, the first formulation of the Swampman problem is not due to Davidson (1987). It was previously hinted at in Millikan (1984) and Papineau (1984):

> Let me put the position starkly- so starkly that the reader may simply close the book! Suppose that by some cosmic accident a collection of molecules formerly in random motion were to coalesce to form your exact physical double. Though possibly that being would be and even would have to be in a state of consciousness exactly like ours, that being would have no ideas, no beliefs, no intentions, no aspirations, no fears no hopes. (Millikan, 1984, p. 93)

> Suppose, for instance, that you didn't exist, but that a being just like you had spontaneously assembled itself a moment ago as a result of some cosmic accident, some random coagulation of just the requisite molecules, and now found itself in just your situation. Wouldn't that being have just the same beliefs, and about just the same objects, as the beliefs you actually have? I do indeed want to deny this. And I recognize that denial is, to say the least, counterintuitive. (Papineau, 1984, p. 565)[44]

While the key ideas of the objection are contained in these quotes, it might be useful to consider Davidson (1987)'s own formulation of the example, since it gives some details that have been important in the discussion.

Davidson considers a similar case, in which a double of him happens to be created by a lightning bolt and, at the same time, he is fulminated:

SWAMPCREATURE Davidson goes hiking in the swamp and is struck and killed by a lightning bolt. At the same time, nearby in the swamp another lightning bolt spontaneously rearranges a bunch of molecules such that, entirely by coincidence, they take on exactly the same form that Davidson's body had at the moment of his death. This being (let us call it 'Swampman') happens to be an exact double of Davidson.

When confronted with this sort of cases, most people have the intuition that, since Swampman is an exact duplicate of Davidson, it must share

44 In fact the idea of Swampman seems to be a very popular example that was used by many philosophers in the mid 1980s. For instance, Block (1986, p. 659) argues: "One problem [with the Fodorian view around 1986] is that one cannot rely on evolution in such a simple way, since one can imagine a molecule-for-molecule duplicate of a baby who comes into being by chance and grows up in the normal way. Such a person would have language with the normal semantic properties, but no evolutionary 'design'."

all its representational states. Hence, we can formulate the following intuitive principle:

SwampDavidson Swampman has the same representational states (beliefs, desires, perceptual states,...) as Davidson.

In a nutshell, the objection consists of two claims: (1) SwampDavidson is extremely plausible (2) teleosemantics is incompatible with SwampDavidson.

First of all, I think (2) is true; indeed, it is quite straightforward that teleosemantics does have this implication. In general, any account that bases content attribution on having a particular kind of history has this consequence, and it is not difficult to see why. Teleosemantics assumes Etiological function. But, according to Etiological function, a mechanism has a function only if its ancestors (past mechanisms from which the actual one is a reproduction) were selected for a certain task. However, *ex hypothesi* the mechanism that produces certain states in Swampman's brain does not have ancestors and hence has not been selected for.[45] For this reason, on a teleosemantic view, Swampman cannot possess representational states. Since the brain states of Swampman have not been selected for, they do not satisfy Selection for; thus, they also fail to fulfill Etiological function, and consequently Third Sender-receiver and Third Content. So Teleosemantics has the consequence that a swampcreature lacks representational states. 'Teleosemantics is forced to say that, since Swampman has no selectional story, he has no contentful beliefs and desires' (Papineau, 2001, p. 281).

The traditional reaction by teleosemanticists has been to bite the bullet and accept that the theory has the counterintuitive result that Swampman does not have representational states (Millikan, 1984, p.93, 1993, 1996; Papineau, 1984, p.565, 1998; Dretske, 1995; Neander; 1995) I agree with this reply. Furthermore, I also think that, even if teleosemanticists must bite the bullet and accept that it is a consequence of the theory that Swampman lacks beliefs, desires or perceptual states, there are many ways of doing this consequence more palatable. The main strategy is to try to argue against (1).

However, before considering in more detail some successful replies, let me discuss two suggested replies that I think are wrong.

3.3.4.1 *Unsuccessful Replies*

impossible swampmen In certain passages, Millikan (1996) seemed to be suggesting that Swampmen are impossible. Her argument is that there might be some features of human beings that are just impossible to replicate in a different being made out of such a different material. Furthermore, she seemed to be assuming that if Swampmen were impossible, they would not constitute a counterexample to teleosemantics. If both claims were true, that would constitute an interesting reply to the Swampman objection.[46]

However, there are two serious worries with this reply. First of all, I doubt it can be shown that the accidental creation of an exact physical

45 Notice that Davidson does not qualify as the Swampman's ancestor. Swampman is not *supposed* to be a copy of Davidson; it is a sheer coincidence that they look exactly the same in all respects.

46 However, it is important to stress that Millikan's official reply is to bite the bullet and accept that Swampman is possible and lacks representational states. Here my main interest is not to discuss whether Millikan in fact endorsed this reply, but only to block such a response.

duplicate of a human being is *nomologically* impossible. It seems that, for all we know about physics, it is nomologically possible (even if extremely improbable) that a physical process that has nothing to do with evolution or reproduction (i.e. a lightning bolt, a freak cosmic accident,...) generates a being that is molecule by molecule an exact double of Davidson. Furthermore, we can imagine alternative processes by means of which Swampman-like creatures can be generated. For instance, Sebastian (2011, p.150-3) describes a 'Zombie Machine', which is composed of a computer connected to a DNA synthesizer. The computer randomly generates sequences of 0s and 1s, which cause the production of molecules by the DNA synthesizer. We could then plug these resulting molecules into a (randomly selected) cell with the basic required proteins. The 'Zombie Machine' would probably generate many different things: dinosaur-like organisms, bacteria-like organisms... Now, some of them might develop into a being that resembles very much human beings. Again, the idea is that the teleosemanticist has to assume that these beings have no intentional states. Needless to say, a Zombie Machine is surely nomologically possible.

Secondly, and more importantly, even if swampcreatures were nomologically impossible, but *metaphysically* possible, they would constitute a counterexample to Teleosemantics. As I argued in chapter 1, teleosemantics aims at lending support to a *metaphysical* supervenience of semantic facts on physical facts (or $\varphi-$facts). So any teleosemanticist is committed to the claim that, if there is some metaphysical possible world where there are swampcreatures, they lack intentional states. If there is a metaphysically possible world such that (1) there is a swampcreature (2) SwampDavidson is true, then teleosemantics is in false.

Finally, even if certain intrinsic properties of human beings make it extremely difficult to create a human being by accident, it seems that similar creatures could pose the same problem: swampfrogs, swampdogs or swampbacteria. It is extremely plausible that such beings could be created by a cosmic accident or Zombie Machines, and according to teleosemantics they would lack representations. Of course, someone might argue that the simpler the organism, the less counterintuitive the objection against teleosemantics is (because it is less clear that a physical copy of a bacteria has to share its representational properties). Nevertheless, a similar objection could be formulated with some organisms that, intuitively, have intentional states, like dogs or frogs.

PAPINEAU (1998) ON SWAMPMAN As I argued, the most common reply to the Swampman objection is simply to bite the bullet and accept that Swamppeople do not have mental representations (Dretske, 1995, Millikan, 1993). Indeed, Papineau (1993) also used to defend this reply, but apparently a graduate student convinced him that he was wrong. As he tells the story, this student pointed out to him the following:

> If [Swamppeople] have not mentality, as teleosemantics implies, then it would seem to follow, absurdly, that it would be right to kill Swampeople and eat them as meat. (Papineau, 2001, p. 281)

According to Papineau, this conclusion is unacceptable. Even if we can accept (with some difficulties) that Swampeople lack representations, it is much harder to hold that we do not have any moral obligations towards them. So Teleosemantics is in trouble.

An obvious reply he considers is that killing Swampmen may be wrong because, even if they lack representational states, they can still be conscious. However, he dismisses this argument on the following grounds:

> No doubt cows and pigs have some kind of conscious sentience, but to most people this does not make it wrong to kill them quickly and painlessly. Killing sentient beings is only clearly wrong when they also have complex enough minds to make plans, form relationships, engage in projects, and so on. (Papineau, 2001, p. 281)

Papineau thinks that we should attribute representational states to Swampeople, because that is the only way of explaining our moral judgments concerning them.

I think that 'the moral argument' (as I will call it) is unsatisfying. On the one hand, it is not obvious to me that it is only clearly wrong to kill beings with complex minds (see Singer, 1975). On the other, Papineau does not consider alternative ethical views in order to justify his claim. For instance, one could endorse a version of rule utilitarianism, according to which it is a moral rule that killing beings that behave like us is wrong. Prima facie, that looks like a plausible rule, since very often beings that behave like us are indeed humans.

In any event, Papineau's recent view is that swampmen can have intentional states. Of course, that puts some pressure on his teleosemantic account of content. In order to resolve this tension, he tries to show that Teleosemantics and SwampDavidson[47] are compatible by reconsidering in what sense teleosemantics is a reductive theory.

According to his new position, teleosemantics does not seek to provide a real definition, in the sense chemistry tries to provide a theoretical definition of water (see 2.1.1), but it only tells us what fills in the 'belief' and 'desire' role *in the actual world*:

> Swamppeople only follows if this essential core is conjoined with the claim that "belief" and "desire" are rigid designators of those states. (...) But it is equally consistent with the central core of Teleosemantics to hold that belief and desire are not rigid designators, and that Swamppeople do have beliefs and desires, on the grounds that in the context of Swampassumptions these psychological terms do not refer to selectional states after all, but to states that would then be present in Swamppeople (p.13)

So according to Papineau, 'belief' and 'desire' designate some roles that in the actual world are realized by states with certain selectional stories but in other worlds can be realized by different entities. 'Belief' and 'desire', then, differ from terms like 'water' and other rigid designators, in the sense that they may pick up different things in different worlds. That is supposed to explain our intuition that Swampmen have beliefs; swampmen instantiate the belief role, but their realizer is different from ours. In the actual world the 'belief' role is satisfied by a set of states with a certain selective story, and in Swampmen-worlds the realizers might be extremely different.

47 Probably, for reasons that will be clear below, Papineau would not accept SwampDavidson as I formulated it. Instead, he would accept some closely related claim, like Swampman or Swampthing (see below). This is irrelevant for the arguments in this section.

BELIEF ROLE First, Papineau holds that the very same belief role might be realized by different states in other possible worlds. The realizer of the belief role in the actual world happens to be a set of states with a certain selective story, but in other possible worlds there are swampmen, which have states with a non-selective story that also instantiate this belief role. Papineau's view is that the fact that Swampmen instantiate this role entails that Swampmen have contentful representations (what, in turn, explains why it is wrong to kill them). On this interpretation, teleosemantics asserts something about the actual world, namely that the belief role is realized by states with certain selective story.

A first problem with this view is apparent: if we assume that certain states in the actual world and states in other possible worlds have the same content in virtue of instantiating a certain belief role, no matter whether their realizers have been selected for or not, then a belief's content is independent of selection story. So, on this interpretation, teleosemantics would not say anything interesting about what content is. Even if teleosemantics would be taken to say something about the actual world, it will not be saying anything philosophically important about the content of beliefs, because this is fixed independently of the selection story.

In other words, if the role of teleosemantics were only to explain what fills the belief role in the actual world, it would lose much of its interest. In that case teleosemantics would not be telling us anything about what beliefs *essentially* are. We would have to abandon the idea that teleosemantics tells us something about the nature of representational states. What makes a state a representational state would be entirely determined by the fact that it instantiates the belief role.

MODALITY Secondly, let us grant for the sake of the argument that (1) human beings in the actual world and Swampmen in other possible worlds instantiate the same belief role (2) these roles are realized by different entities in the two possible worlds (3) the fact that Swampmen instantiate the belief role explains why they have contentful representational states. Still, it seems that this objection fails to solve the Swampman problem because (as we saw when discussing one of Millikan's replies), Swamppeople are not only metaphysically possible, but also nomologically possible. So, since Swamppeople can exist in the actual world and (and according to Papineau) teleosemantics tries to analyze what are the realizers of the belief role in our world, the possibility of Swamppeople renders teleosemantics false, even if the truth of teleosemantics were restricted to realizers of the belief role in the actual world.

In short, I think Papineau's recent view of swampcases and teleosemantics is insufficiently motivated, since there are other plausible ways of explaining why we feel obliged to behave morally towards Swampeople. Furthermore, I think his recent interpretation of teleosemantics would make it completely uninteresting.

Let us move on to the replies that are more likely to succeed.

3.3.4.2 *Reply*

THEORY *a posteriori* I think that the strongest argument in favor of the teleosemanticist is that naturalistic theories such as teleosemantics are supposed to provide a real definition of a certain phenomenon, as

defined in chapter 1 (see 1.1.1) and chapter 2 (see 2.1.1). In other words, naturalistic theories are *a posteriori* theories of representation (Papineau, 1996). They are supposed to provide an account compatible with the evidence gathered by science and such that it fits as much as possible our pre-theoretical intuitions. However, it is likely that some of the claims that derive from a philosophical investigation counter our pre-theoretical intuitions. Swampman may be such a case. In this situation, given the overwhelming number of cases where teleosemantics gets it right and taking into account the exceptionality of swampcases, it is reasonable to dismiss swampmen intuitions (Millikan, 1984; Papineau, 1996).

Consequently, whereas I concede that the intuitiveness of SwampDavidson tells against teleological theories of representation, the a posteriori nature of any naturalistic enterprise enables us to dismiss this countervailing intuition.

Nonetheless, I admit that swampcases constitute an unwelcome consequence of the theory. So, even though I think the a posteriori status of teleosemantics shows that we should not abandon the theory just because of swampcases (assuming they are extremely rare), I think we should try to make this consequence more palatable. Fortunately, I think there are independent arguments suggesting that swampcases are a poor guide to the nature of content and representation. That is what I will intend to show in the remainder.

SWAMPDAVIDSON, SWAMPMAN AND SWAMPTHING My argument against swampman has two steps. First, I will show that SwampDavidson is incompatible with any reasonable view of content, biological function and trait individuation. So, either one accepts SwampDavidson or else one accepts a set of widely held views in philosophy and biology. After presenting these plausible principles, I hope it is obvious to most people that the second option is preferable. We should not give up our views on these topics (supported by extensive and well-established arguments), just because of the intuition supporting SwampDavidson.

Secondly, I will show that there is a way of cashing out the swampman intuition that respects these common philosophical and biological views. However (and this is the second step of my argument), if Swampman is spelled out in a way compatible with reasonable views of content, biological function and trait individuation, then Swampcases lose much of its intuitive force. So, I tentatively conclude, Swampcases have only some intuitive force when they are cashed out in a way that clearly threatens some of our strongest views in philosophy and biology.

Let us begin by pointing out that there are certain thoughts that very probably SwampDavidson cannot have. For instance, it is usually thought that in order to have singular thoughts there must be a causal or primitive relation of some sort between the thought and what the thought is about (Evans, 1982; Campbell, 2002; Jeshion, 2010). However, whatever causal or primitive relation is needed for a thought to be a singular thought about Aristotle or Davidson's Mother, this relation is surely missing between SwampDavidson and them (Levine,1996). So, if that is right, SwampDavidson cannot have singular thoughts about Aristotle or Davidson's mother. Furthermore, since SwampDavidson has just been created, he cannot remember when he was young, while

Davidson clearly could. Nor can he remember Davidson's youth, since a plausible requirement for a subject to remember X is that X figures in the causal antecedent of that memory or, at least, that a past perception or thought about X has caused this memory (Schellenberg, 2010). Nothing like Davidson's past has caused SwampDavidson's current brain states. So SwampDavidson and Davidson also differ in their memories. Thirdly, if externalism about thought is true (and assuming that Fodor's Asymmetric Dependence Theory is wrong, as I argued in 1.2.4- see also Antony, 1996; Levine, 1996), SwampDavidson cannot think of water or cows. Whatever causal relation is required for a subject to think about natural kinds is surely missing between water or cows and SwampDavidson (Millikan, 1996). In fact, if some form of social externalism is true (Burge,2007), SwampDavidson cannot think of arthritis, Smallpox, or many other items either. I think these claims are, if not completely uncontroversial, at least strongly compelling.

Notice that all these assertions clash with SwampDavidson. The intuition elicited by SwampDavidson is that it can have the same beliefs as Davidson because it happens to be an exact double of him. But most philosophers are willing to claim that, concerning beliefs about water, past events and singular thoughts (to mention just a few), SwampDavidson can not have them. Hence, most people are committed to denying SwampDavidson. SwampDavidson cannot have the same representational capacities as Davidson.

Obviously, this is not yet a rejection of the whole argument, since one could come up with a different principle that is compatible with all these claims about content and still has a similar intuitive force as SwampDavidson. For instance, one could endorse Swampman:

Swampman: SwampDavidson has some representational states (beliefs, desires, perceptual states..), whose content differ from Davidson's.

Even if Swampman is very similar to SwampDavidson, notice that in the transition from SwampDavidson to Swampman the intuitive force of this counterexample is partially lost. Swampcases lend support to SwampDavidson; but some arguments strongly suggest that SwampDavidson should be rejected. We have good reasons for denying that Swampcreatures have exactly the same representational powers as their duplicates. That is why we formulated Swampman (which contradicts SwampDavidson). Swampman still has some intuitive appeal, but it is a significant result that we have to reject the strongest intuition against teleosemantics because of its preposterous consequences.

Of course, someone might argue that Swampman is still very intuitive. One thing is to claim that Swampman cannot have beliefs about Davidson's past or about Smallpox, and another that it cannot have beliefs. After all (the argument runs) Swampman has a neuronal structure identical to Davidson's, so even if one is externalist and assumes that there are certain things Swampman cannot think about, Swampman must be thinking something when its neuronal states are firing.

This is probably the sort of nuanced objection most people have in mind (rather than something as strong as SwampDavidson; but see Antony (1996)). I take it that even if most people are externalists and hence have to deny that our intuitions concerning swampcases are hundred per cent reliable, the claim that its brain states are representing something seems undeniable.

Now, the next important issue is: does Swampman in fact have neurons or a brain? Let us move first to a closely related question: is Swampman a *man*? Since Gishelin (1974) and Hull (1978), I think few would accept that Swampman is a member of the human species. In general, it is commonly accepted in biology that species cannot be individuated by a set of necessary and sufficient conditions all members of a species share (Wilson, 1999).[48] The most plausible candidate, DNA, seems to fall short of dividing species in a satisfactory way. For one thing, there is almost no gene in the human DNA that lacks an allele (so there is not DNA structure that is necessary for an organism to be a member of *homo sapiens*). For another, if we look at the evolutionary past, there are many organisms that we want to classify as belonging to the same species but which have a very different DNA. Furthermore, a single cell roughly contains the same DNA as a whole human being, but it is not a human (putting many cells together does not solve the problem). Biological species are usually individuated by appealing to causal connections between members of the species (Sterelny and Griffiths, 1999, p.183; Boyd, 1999a; de Queiroz, 1999; Ereshefsky, 2010). In other words, species form first-order reproductively established families, in the sense of FIRST-ORDER REF defined above (indeed, they are also Darwinian populations). A necessary condition for an organism to be a human is that it has been produced by a human. Therefore, Swampman is not a human.

Similarly, it seems that Swampman does not have a brain, or a heart, or kidneys. Think about Swampman's *heart* (let's call it 'swampheart'). Is swampheart really a heart? One might argue that since SwampDavidson's swampheart exactly resembles Davidsons' heart, it must also be a heart. The problem, however, is that, if swampheart is a heart, then *being a heart* is not a reproductively established family (in opposition to FIRST-ORDER REF, see 3.2.4), since there is an item that belongs to that family without being a reproduction of a past heart.[49] But, again, that seems to be in tension with the way traits are individuated in biology (Neander, 1996; 2002). A malformed kidney is a kidney because it has been caused by a past kidney (or caused by a gene that encodes for a kidney). In contrast, swampkidney, swampheart or swampbrain is not a copy of any other kidney, heart or brain. So Swampman does not have any kidney, heart or brain (Neander, 1991, p.180). Similarly, Swampman does not have neurons.

Indeed, I think that there is a different way of showing that Swampman cannot have a heart or neurons. As I argued, if Swampman had a brain and neurons, we would be committed to reject REPRODUCTIVELY ESTABLISHED FAMILY. But then, on the plausible assumption that having a certain function is an essential property of neurons and hearts, we will also have to give up SELECTION FOR and ETIOLOGICAL FUNCTION. In particular, if swamphearts were hearts and hence had the function of pumping blood, we would be forced to abandon the view that having a particular history is required for a trait to have a function. But, as I extensively argued earlier (see 2.1.2) that yields an utterly unappealing

48 Unless we include causal and historical properties within the necessary and sufficient conditions, of course.

49 Remember that swampheart is not a copy of Davidson's heart, but *happens* to resemble Davidson's heart by a freak accident. As we saw, items that belong to the same Reproductively Established Family must be causally connected. The proof that this causal connection is missing between Swamphearts and hearts is that certain counterfactuals fail to hold here: even if Davidson's heart had been radically different, the swampheart would have been exactly alike.

theory of function. If we accept that swamphearts have a function, in virtue of what can a swampheart malfunction? We saw that resorting to what most hearts actually do yields counterintuitive results (remember Neander's dictum: one cannot health a disease by spreading it around). Similarly, how can we distinguish the accidental from the functional effects of Swamptraits? Is a function of the swampnose to support glasses? Why not? I think that rejecting the historical requirement of ETIOLOGICAL FUNCTION yields a reductio of any theory of function. Since ETIOLOGICAL FUNCTION is the best account of function we have and functions are essential properties of traits, we should deny that Swampman has a brain, neurons or lungs.

The same argument can be formulated directly. I showed earlier (section 2.1.1) that, in order to make sense of some of the properties of functions (distinction between functional/accidental effects and the possibility of malfunction), a necessary condition for a trait to have a function is that it has an adequate selective story. But if this is required, swamphearts have no function. And if swamphearts have no function, they are not hearts. So swamphearts are not hearts.

So I have shown that Swampman is not really a man, and that none of its organs are of the same type as the ones had by Davidson. As a result, it follows that in fact, SWAMPMAN should be formulated more accurately in the following terms:

SWAMPTHING SwampDavidson is not a human and has no brain, no neurons, no heart, no kidneys,... in fact, none of its swamporgans has any function whatsoever. Nonetheless, he has representational states (beliefs, desires, perceptual states,...), whose content is different from Davidson's.

SWAMPTHING presents the only claim that is compatible with externalism, trait individuation as it is carried out in biology and the only notion of function that has any plausibility. Of course, the problem with SWAMPTHING is that it has lost most of its intuitive force. SWAMPDAVIDSON was very intuitive but very implausible; SWAMPTHING is coherent with our view on a wide range of matters, but I think not very intuitive anymore. Of course, I grant that SWAMPTHING has *some* intuitive force and that it is incompatible with teleosemantics (as formulated above). But the key point I wanted to stress is that a sensible formulation of swampcases are not as counterintuitive as might seem at first glance. I think that makes the bullet much easier to swallow.

3.4 CONCLUSION

In this last chapter of the first part of the dissertation, I have formulated the teleosemantic framework in a way that can deal with most objections and I have compared THIRD TELEOSEMANTICS to several alternative views on the nature of function and representation. It is now time to put this framework to work and see how it bears on the complex representational capacities of humans and other organisms. This is the task of the second part of the dissertation.

Part II

PERCEPTION AND COGNITION

PERCEPTION AND COGNITION

In the first part of the dissertation I devised a set of tools that should enable us to provide a naturalistic account of the general phenomenon of representation. My main goal was to show that semantic properties could be analyzed in non-semantic terms. In particular, I showed that there is a plausible way of reducing intentionality to φ-facts, that is, to facts that probably metaphysically supervene on physical facts (see 1.1.1). I argued at length that the theory I offered can overcome the problems that affect other proposals and that it seems to yield the right results in some cases.

However, devising such a toolkit was only part of the task of this dissertation. The original goal was not only to establish that such a reduction is possible *in principle*, but also to show how a particular set of intentional states (which, as I argued, have a privileged status) can actually be reduced: perceptual and conceptual states. As I said in chapter 1, these intentional states are of special relevance for two main reasons. On the one hand, they are the kind of representational states that more stubbornly have resisted any attempt of reduction. On the other, if perceptual and conceptual states can be naturalized, then the idea that the rest of intentional states (linguistic expressions, artifacts,..) can be also reduced becomes extremely plausible.

Accordingly, in this second half of the thesis I would like to show how the framework set up in the first half and condensed in Third Teleosemantics can be applied to some central human cognitive abilities. The project of the second part of the thesis, thus, is to focus our attention on some cognitive mechanisms and states and show how these structures can be naturalized. As a consequence, this second half is going to be much more empirically oriented. We will also pay attention to the way cognitive scientists proceed and attribute representational content to certain state, and see whether it can be accommodated in our framework.

With this goal in mind, in chapter 4 I concentrate on perceptual abilities and in the last two chapters on concepts.

4 NEUROSEMANTICS

The main aim of this first chapter is to show how the teleosemantic structures we defined in part I are instantiated in perception. I will argue that it is very plausible that perceptual states indeed satisfy the conditions set up in Third Teleosemantics for being representational states. That project requires a naturalistic account of the semantic properties of simple and complex neuronal representations, what some people call *neurosemantics* (Eliasmith, 2000; Mandik, 2003; Ryder, 2009). This is a field that remains largely unexplored by naturalistic theories.

This chapter is organized in three main sections. First of all, I will motivate my approach and put forward some preliminary problems concerning the application of the teleosemantic ideas developed in part I of this dissertation to neuronal systems. Then, I will describe in some detail how teleosemantics is supposed to apply to the toad's perceptual system and the human visual system. This discussion will enable us to provide an analysis of a central cognitive ability that is going to be crucial in the following chapters: perceptual tracking. Finally, I will address the relationships between the theory I am defending and certain issues in the philosophy of perception such as the debate on non-conceptual content or the question whether the content of perceptual states is singular or general.

Let us start, then, by framing a neuroteleosemantic approach to cognitive states.

4.1 TELEOSEMANTICS AND PERCEPTION

4.1.1 *Motivations and Prospects of Neurosemantics*

First of all, let me try to argue why we need a neurosemantic account. i.e. a naturalistic account of the semantic properties of complex neuronal structures (with special focus on perceptual mechanisms). There are at least two general motivations for a naturalistic account of neuronal states.

On the one hand, this topic is obviously significant and important in its own right. One of the many questions in philosophy (and neuroscience) is whether states in the brain represent certain features in the environment, and if so, what process determines this representational status (see below for some references). This discussion is specially important in the context of perception; some of the most hotly debated issues are whether perceptual states are endowed with representational content and how should we conceive these contents. Thus, in developing neurosemantics, we are making an important contribution to several ongoing debates in philosophy and science.

A second related motivation concerns the main project of this dissertation: to develop a naturalistic account of concepts and higher-order cognitive abilities. As we will see in the next chapter, one of the major reasons why current naturalistic theories of concepts fail is precisely because they do not provide a naturalistic account of perceptual content. And, as I will argue, perceptual content can only be naturalized if one

has previously developed a neurosemantic approach (see 6.3.2). One of the chief goals of this chapter is to set the ground for a solution to this common pitfall.

Those are, I think, two remarkable motivations for undertaking a research on a naturalistic theory of neuronal representations. A different question, however, is whether a naturalistic account along the lines of teleosemantics is likely to succeed. Why should we think teleosemantics has any chance of being true in the context of neuroscience? There are three nice features of the theory that should prompt an optimistic attitude towards the neuroteleosemantic project:

- First, neuroscience is pervaded with representational talk (Eliasmith, 2000; Kandel et al., 2000). It is extremely common among neuroscientists to claim that certain brain structures represent (or detect) particular features, but this notion of 'representation' is usually left unexplained. For example, it is standardly claimed that in the primary visual area there is a retinotopic representation of the visual field or that a pattern of neuronal activation in a certain brain region corresponds to an 'edge detector' (see 4.2.2). Furthermore, it is assumed that this notion of representation is shared with other areas of cognitive science (Sternberg, 2009). Hence, it is reasonable to suppose that the representations attributed by neuroscientists belong to the same kind of representations we have been talking about all along (possible arguments to the contrary will be discussed in 4.1.3).

- Secondly, it is common to describe patterns of activation in certain neuronal structures not only as representational states, but also as representing distal features. For instance, if we focus on the visual system, it is very common to talk about detectors of motion, color, shape or size, even if the input employed by the brain in order to perform all its computational tasks is almost exclusively composed of proximal cues: photons impinging the retina, proprioceptive information about eye position, waves altering the auditory system, and so on (we will discuss some examples). The distality of content is a significant aspect of current neuroscientific explanations because we saw that a pattern of covariation or statistical dependence cannot account for the representation of distal features. I argued in 2.3.3 that only teleosemantics can make sense of the fact that some states represent distal entities, due to its appeal to the needs of consumers. That suggests that teleosemantics might also be the right neurosemantic theory.

- A further aspect that indicates that a teleosemantic account is the right place to look for a naturalistic neurosemantic account is the following: a common assumption in neuroscience is that *use* of a particular mental state determines what it represents (Bertenthal, 1996, p. 415-16; Bechtel, 1998, p. 337). In other words: if the firing rate of a certain brain structure is *used* in order to compute a certain feature, it is commonly assumed that this consumption will reveal the representational content of the state (Eliasmith, 2000). The idea that use determines content is a striking feature shared by teleosemantics and neuroscience.

Therefore, there are enough good reasons for thinking that neuroteleosemantics is a promising project worth exploring in detail.

4.1.2 *Preliminary questions*

Despite the clear motivations set up in the previous section, a remarkable (and surprising) fact is that there have been few attempts to carry out a naturalistic account of neuronal representations. Generally, the leading examples motivating naturalistic theories involve extremely simple and automatic mechanisms in cognitively unsophisticated organisms: a sender that is activated by a certain cue and a consumer that reacts towards this very same cue by behaving in a certain way. Frogs (the organism that has attracted an incredible amount of attention in the literature) surely possess a quite complex brain which performs many different computations before performing any behavior, but all these complexities were abstracted away when considering how teleosemantics applies to them. As the example has usually been described (for instance, in Fodor, 1990; Agar, 1993 Sterelny, 1990) the process is supposed to be extremely simple: a black shadow moving around, a mental state going on, and a certain behavior towards a fly. Nothing like a complex neuronal structure or cognition was being described or explained by these examples.

Of course, this oversimplification was justified when the goal was to get clear about the basic teleosemantic framework (but see Papineau (1998) and Neander (2006) for some criticisms). Nevertheless, after so many years of promising naturalistic proposals, one should be surprised to see that few people have attempted to develop a more detailed *neuroteleosemantic* proposal. Recent proposals in that direction include Eliasmith (2000), who has developed a theory based on statistical dependence (along the lines of Relative Indication), Ryder (2006, 2009) and Cao (2012), who have assumed a broad teleosemantic framework. I will show in due time to what extent my account differs from theirs.

There are, however, some reasons that may help to explain why there have been too few neuroteleosemantic proposals. Once we move from simple representational systems to cognitive systems, there are at least three aspects that become highly problematic. It is not unreasonable to think that these difficulties have discouraged most philosophers to pursue this line of research. So let me first describe in some detail the specific problems that concern cognitive capacities and then show how a naturalistic account can overcome them.

I think one can distinguish three important challenges of any neuroteleosemantic account:

genuine representations? A popular view in philosophy holds that representational talk in the context of many neuronal structures such as sub-personal states[1] is a mere as-if way of talking. In other words, some people have argued that many sub-personal states in the brain (such as brain states in early perceptual processing) are not full-blown representations. They certainly accept that there is some correlation or covariation between states in the brain and certain external features, but reject the view that, strictly speaking, these sub-personal states should qualify as representational (Burge, 2010; Cao, 2012; McDowell, 1994). These people hold that there is a gap between

1 How to properly define 'sub-personal state' is a very controversial topic I cannot get into here. In this discussion, the sub-personal states I have in mind are internal states of an organism that may not be phenomenally conscious or consciously accessed (in the sense of Block, 1997a), but which play important computational roles. Neuronal activation in the Lateral Geniculate Nucleus or the Hippocampus are two examples.

states that merely correlate or indicate states of affairs, which are located at the level of sub-personal processing, and full-blown representational states, like full perceptual experiences and thoughts. If that were true, then the project of providing a teleosemantic account of sub-personal states of the brain that also applied to perceptual and conceptual representations would be doomed.

SENDER-RECEIVER SYSTEMS IN THE BRAIN? A more specific problem with developing a neuroteleosemantic approach is that, in cognition, the task of identifying producer and consumer systems becomes extremely complicated. In that respect, notice that the most popular cases in the teleosemantic literature clearly instantiate a Sender-Receiver framework, since the producer and consumer systems are usually identified with some well-established mechanisms like the snapping system, the visual system, the motor system or the digestive system. However, once we move into the brain, identifying systems and subsystems becomes much more difficult (Cao, 2012; Godfrey-Smith, personal communication). For instance, only a very small subset of sub-personal systems are activated by external cues; most of them are produced by some internal activity. Similarly, all but a very small group of consumer systems have effects on other internal systems of the organism. In other words, most representational systems in cognition receive certain cues from previous cognitive systems and produce states that will primarily have effects on another representational systems. A more precise way of expressing this idea is that most neurons are not *sensory* or *motor neurons*, but *interneurons* (Kandel, 2000). As Cao (2012, p. 50) claims, "The world of the neuron (i.e. the world in which it is competent to take action) consists entirely of more neurons and the supporting cells around them". How could a sender-receiver model be instantiated within a complex chain of neuronal states? The fact that in cognition behavior is always mediated by a myriad of steps and computations, has made it difficult for many people to see how the teleosemantic framework could be implemented in sub-personal systems.

DECOUPLED REPRESENTATIONS Finally, it seems that there are some cognitive states that are not supposed to elicit any particular behavioral response. Organisms with the capacity of entertaining what Sterelny calls 'decoupled representations' can have states which represent certain states of affairs but need not react in any particular way to them. Decoupled representations are 'internal cognitive states which (a) function to track features of the environment, and (b) are not tightly coupled functionally to specific types of response' (Sterelny, 2003, p. 31). For instance, we can have a perceptual representation of a certain scene without there being (in Normal conditions) any behavior or even intention of acting in a certain way. However, according to the teleosemantic framework I suggested in part I, content is utterly determined by the activities of the consumer system. How can perceptual states be endowed with content if there is no behavior that Normally ensues them? And, more generally, how can we account for the existence of decoupled representations? This is what Matthen (2006, p. 150) calls the 'Problem of Multiple Responses'.

These are serious challenges to any naturalistic theory of the content of neuronal states, and a big part of this chapter is devoted to overcome

these anxieties. In the following two sections I will directly address objection 1 and (partially) 2, which I think require an independent discussion. I will tackle the third worry after sketching the main ideas of a neuroteleosemantic account.

After addressing these general issues, I will present in detail three real examples. First of all, I will indicate how the teleosemantic framework can be implemented within a (very simple) computational structure, using Mandik's AI model (which roughly approximates certain real cases). Secondly, I will focus on the toad's cognitive abilities, in order to show how the sender-receiver model can be instantiated in relatively sophisticated brains. Finally, I will move to the human visual system and I will arguethat these difficulties can also be met there.

4.1.3 *First difficulty: Genuine representations in the brain?*

Let us start by considering the first worry raised in the previous section: can states in sub-personal structures qualify as full-blown representations? I would like to argue for an affirmative answer.

First, it is noteworthy that while many philosophers that read cognitive scientists think philosophical theories should try to be as much in accordance as possible with cognitive science, neurosemantics is a place where this assumption seems to be abandoned. As a matter of fact, most cognitive scientists assume and assert that sub-personal states such as a pattern of activity of a set of neurons in a given brain area might represent a certain environmental feature. As Ryder (2009, p. 18) claims:

> In the neuroscientific literature, the term 'representation' is often used just in case there is a neural 'detector' of some type of environmental stimulus, which need not be particularly unruly. For instance, ganglion cells in the retina are sometimes said to 'represent' the impingement of multiple photons in a single location

Similar remarks can be found in Kandel (2000) and Eliasmith (2000). However, a large number of philosophers in many different areas of research think that representational talk at this stage is just a loose way of talking. For example, Burge (2010), Pylyshyn (2003), Raftopoulos (2009a, ch. 4-5), McDowell (1994) or Cao (2012) claim that sub-personal states in perceptual processing should not qualify as full-blown representations. They usually claim such states *indicate* certain features or *carry information*, but should not be considered as representations in the sense defined in Part I, that is, as states with veridicality conditions.

For instance, Cao (2012) has recently argued:

> Contrary to received views, neurons will have little or no access to semantic information (though their patterns of activity may carry plenty of quantitative, correlational information) about the world outside the organism. Genuine representation of the world requires an organism-level receiver of semantic information, to which any particular set of neurons makes only a small contribution. (Cao, 2012, p. 49)

It is worth stressing that this is not a mere terminological quibble; whether some states qualify or not as full-blown representations is

an important metaphysical question that will surely have decisive consequences in many fields. For instance, if sub-personal states are representations in the full sense, they possess veridicality conditions, and hence certain normative properties that they would otherwise lack. Secondly, in some scientific enterprises states might be classified in different ways depending on the answer to that question (Burge,2010, see below). Thirdly, as I will argue in this second part of the thesis, only if sub-personal states qualify as proper representations can they play a pivotal role in the naturalization of content. In other words: if sub-personal states cannot be naturalized with Third Teleosemantics, then it is very unlikely that more complex representational states will be naturalizable. Therefore, this is not primarily a discussion about words: decisive issues hang on the representational status of sub-personal states.

My goal in this section is to defend the use of representational talk in cognitive science. I hold that many sub-personal states (like neural firings in early perceptual processing) are representational in exactly the same way thoughts and experiences are representational: they have veridicality conditions and hence can be assessed for truth/accuracy and falsity/inaccuracy. As Millikan (2004, p. 158) wrote:

> To suggest that genuine intentionality, genuine aboutness, with the possibility of misrepresentation, actually occurs at this level may at first seem far-fetched. But the idea idea is that there is intentionality in the sort of way zero is a number. These are the most humble sorts of limiting cases of intentionality. By treating such simple signals as intentional signs, just as by treating zero as a number, we will be able to examine their relations to various successors, and see the continuity between them and their more sophisticated relatives.

Now, since cognitive scientists indeed use the term 'representation' and, furthermore, seem to be assuming that this notion has normative import (after all, they usually talk of *mal*functioning and *mis*representing[2]), any view that takes cognitive science seriously (and, of course, all contenders in this debate surely do) should have strong arguments for modifying the scientist's perspective. Consequently, I think the view that many sub-personal states are representational should be the default position; the onus of the proof is on its critics.

In what follows, I will discuss some reasons for holding that only sophisticated cognitive states like perceptual experiences or thoughts should qualify as representations. In that respect, I have been able to identify two common arguments in favor of the claim that many sub-personal states such as states in early visual processing are not representational: the first is based on explanation and the second on psychological kinds.

EXPLANATION Here is a quote from a recent and influential book, in which it is suggested that most sub-personal states should not qualify as representational states:

> In the cases of some sensory states—non-perceptual ones—saying that the states have veridicality conditions would add noth-

2 Thanks to Miguel Sebastian for that remark.

> ing explanatory to what is known about discriminative sensitivity and the biological function of the sensitivity. Invocation of veridicality conditions and perceptual perspective does not figure integrally in any explanation of these states. Veridicality conditions can be imposed. But invoking them gains no empirical traction, yields no empirical illumination. In such cases, there is no reason to believe that there are representational states. (Burge,2010, p. 395)[3]

This is an extremely common argument both in philosophy and some parts of science (see, for instance, Bermudez, 2003, p. 6-9; Sterelny, 2003; Pylyshyn, 2003, Rescorla, 2013). Again, notice that this debate has important metaphysical implications. The question is not only whether a certain state should be described as being a representation, but whether this state has veridicality conditions, whether it can be true (accurate) or false (inaccurate).

The argument suggested in Burge's quote is based on two main claims. The first one says that if we are able to explain a given state or mechanism in non-representational terms, an attribution of representations is unwarranted. For instance, if a complete explanation of a bacteria can be provided, which only appeals to causal connections, functions, and so on, then we should not ascribe representations to these states (see Sterelny,2003, ch. 2).

The second key claim is that, if we focus on the human brain, there is a clear contrast between cognitive states such as belief, thought and full perceptual states on the one hand and sub-personal states on the other. Whereas in an explanation of the former we can not dispense with representational talk, the latter can be fully accounted for merely in terms of non-representational notions, such as discriminative capacities, biological functions and the like. The idea, then, is that ascribing representations to certain sub-personal states is unjustified because the same phenomenon can be explained in simpler or more fundamental terms. In contrast, when we move to more sophisticated forms of representation (perception, concepts, etc...) reduction to mere causal, informational or historical relations is utterly impossible. Hence, by restricting the application of representational talk to this latter domain, we use representational talk only where it is necessary and informative.

Putting the two ideas together, here is my reconstruction of this common argument:

1. An attribution of representations to x is unwarranted iff a full explanation of x can be provided in non-representational terms.

2. A full explanation of sensory states can be provided in non-representational terms (discriminatory capacities, biological function,...).

3. A full explanation of cognitive states (thoughts, percepts,..) can not be provided in non-representational terms.

- Therefore,

3 We should distinguish this argument from the worry of whether there is any explanatory gain in attributing semantic properties *in general* (Stich, 1983). It is not obvious in which sense semantic properties are explanatory (see 3.3.2.1), but this is not the problem we are trying to address. We are concerned with the question of whether sub-personal states have *the same kind* of semantic properties as beliefs, desires and perceptual states. In which sense all those attributions are explanatory is a different question.

- An attribution of representations to sensory states is unjustified (*by 1 and 2*)
- An attribution of representations to cognitive states is justified (*by 1 and 3*)

I think that the argument is unsuccessful for several reasons.

First of all, consider premise 1, which relies on the idea that if x can be explained in non-representational terms, then an attribution of representations to x is unwarranted. In general, it does not seem to be true that the fact that you can explain x without appealing to a property F shows it is wrong to attribute F to x. For instance, even if the Sagrada Familia has the property of being my favorite's building, I one can probably fully explain the Sagrada Familia (structure, history, outlook,..) without mentioning this property. Similarly, if heat reduces to mean kinetic energy, perhaps one can fully explain the phenomenon of boiling water in terms of kinetic energy. However, that does not mean that the claim 'x is hot' is false or unwarranted. Consequently, the first consideration concerning premise 1 is that, in general, the fact that you can explain a certain phenomenon without appealing to F does not show that a phenomenon lacks F.

A related point is that premise 1 assumes a dubious link between explanation and metaphysics. In principle, it seems that a state qualifies as an F if it satisfies the sufficient conditions for being an F. So whether a state qualifies as a representation primarily depends on what representations are, rather than on whether we need it in order to explain a phenomenon. It is not obvious why the fact that we gain empirical illumination or the fact that a causal-historical account of content is 'so complicated that pragmatically we have no way of telling it' (Cao, 2012, p. 55) should be relevant concerning the metaphysical question of whether a given entity is a representation.

Indeed, if some naturalistic account of the mind is right, then premise 1 is not only prima facie dubious, but blatantly false. Naturalistic accounts aim at providing an explanation of representational phenomena in non-representational terms. Consequently, if any of these reductive theories succeeds, then you should expect that, for any representational state x, there is a complete explanation of x that does not appeal to representations. In other words, if naturalism about semantic properties is true, then there is a full explanation of any representational system in non-representational terms. So premise 1 would be clearly false.

Similarly, I think premises 2 and 3 are problematic, although this point is trickier. The main objection to premise 2 and 3 is the following: if some naturalistic theory of content is true, then there is no contrast between the explanatory power of representational notions in the context of sensory systems and the explanatory power of the same notions in the context of cognitive states. Let me slightly elaborate on this point.

As we saw in 3.3.2.1, the explanatory power of semantic notions is a disputed topic. But it seems that, if they have some explanatory import, it should be the same when they are attributed to sensory systems or cognitive systems. Here is a reason: according to naturalism, all truths about representational states supervene on truths about non-representational states. So, once you fix the truths about biological function, discriminative capacities and so on, all truths about the semantic properties of states are fixed. Whether that shows that the appeal to representations is not really explanatory is unclear, but

what it does indicate is that if semantic notions are explanatory, they should be equally valuable for *all* representations. So if we forget about epistemological worries (i.e. about the relative simplicity of sensory systems in relation to cognitive systems), there does not seem to be any metaphysical ground for distinguishing the explanatory import of ascribing representational states to sensory states or cognitive states.

So I think that the argument against the application of representational notions to sub-personal states stated above is unsatisfactory. Nevertheless, before moving forward, let me make two points.

First of all, one can be concerned about the dialectical situation. On the one hand one could reply that in my arguments I have been assuming that some naturalistic account of the mind is true, and that just begs the question against the non-representationalist. However, that complaint would miss the role this discussion plays into the whole picture. We started this section asking whether there is any principled reason for thinking sub-personal states are not representations, in order to prepare the ground for a teleosemantic account. What I have shown is that a common argument against the existence of representations must assume that naturalistic accounts of the mind fail. If this assumption is not made, then the premises of the argument are jeopardized. Given that in the previous chapter I already addressed the most important objections to the teleosemantic project (see 3.3.1) and given the good prospects of a teleosemantic account of neural states (see 4.1.2), I think it would be unreasonable to start off this discussion by already assuming the failure of the project.

Similarly, one could object against the general dialectical strategy that, whereas it follows from my particular view on representations that sub-personal states such as states in early visual processing qualify as such, Sterelny, Pylyshyn and Burge's favor a different account of what representations are that implies that these states are not representational. How are we to resolve this issue? In that respect, an important point in my favor is that these authors heavily rely on the argument of explanatory idleness in order to argue for their restricted view on representational phenomena. It is primarily by assuming that representational talk is explanatorily vacuous in the case of simple states, that they conclude that 'representation' must be referring to more sophisticated kinds of states. Instead, I provided a whole set of independent arguments in favor of a certain understanding of representational facts (developed in part I). Furthermore, scientists usually talk as if these kind of states were also representational. So, if (as I will argue in this chapter) it follows from this approach that states in early visual processing are representations in the full sense, I think we have good reasons for endorsing this view.

Indeed, in this dissertation I have been presenting and defending a framework that specifies in a principled way what kind of conditions must hold for a state to qualify as a representation (i.e. it provides what I called a 'metasemantic' account of representation). Now, if (as we will see) it follows from the view defended here that many sub-personal states are representational (because they satisfy the conditions set up in Third Teleosemantics) we will be justified in assuming that they are representations. So, if the thesis developed in this dissertation is right, any phenomenon of representation can be reduced to 'weightings of registrations of such stimulation from different bodily sensors, capacities for adaptation or conditioning, neural pathways, and [cru-

cially] biological functions of the system' (see Burge's quote above). If I am right, perceptual and conceptual representations are nothing more nor less than certain states which derive from a a particular group of functional mechanisms.

PSYCHOLOGICAL KINDS The second sort of objection is based on the classification of psychological states into kinds. Burge, among others, complains that if we assume a notion of representation that applies to automatic sub-personal states and the internal states of cognitively unsophisticated organisms, we will fail to capture the important sense in which perceptual states differ from the activation of ganglion cells in the retina and states in simple organisms (similar arguments can be found in Sterelny, 1995).

> A few philosophers and scientists have stretched or deflated representational notions so far as to claim that everything represents something or other. Tree rings represent age, smoke represents fire; the earth's orbit represents the gravitational powers of the sun; and so on. (...) More specifically, these conceptions tend to miss a distinctively psychological kind that constitutively and non-trivially involves perspective and conditions of accuracy. (Burge, 2010, p.27)

Two things should be said here. First of all, I take Burge and others to be arguing against naturalistic theories of content like teleosemantics. However, teleosemantics does not imply that trees represent age or smoke fire. Indeed, none of these things represent anything because there is no sender or receiver that have been designed in order to produce and consume these states (see 2.2.4).

Secondly, it is a platitude that there are striking differences between the bacteria's internal magnetosome, ganglion cells and perceptual experiences. Such differences exist and Burge is right in pointing them out. However, teleosemantics has the resources for explaining this diversity: the remarkable differences between these states can be explained by the diversity of representational contents (distality, variability,..), a difference in the vehicles of representation, in the combinatorial capacity of these vehicles, etc... Hence, there is no need for an explanation of these differences in terms of essentially different kinds of representation, rather than different kinds of vehicles, representational content or combinatorial capacities. Whereas these states are all different (and might be classified as belonging to different kinds by some lights) I think there is an important sense in which all of them are representational states. What I am arguing is that this is the sense that justifies the claim that all of them have veridicality conditions. At least, this is what this dissertation is arguing for.

In conclusion, I think that we lack convincing arguments against the usual assumption in cognitive science that many sub-personal states are representational. As a result, I think that at this point we should side with science and assume that many sub-personal states such as those in early perceptual processing are representational states in a substantive sense. These states have veridicality conditions in the same sense concepts, thoughts or visual percepts do have them (while this does not imply, of course, that there are no important differences between them). Furthermore, if am right, all of them are representations in virtue of

being the products of some mechanisms with biological functions. The fact that I have independent reasons for assuming that these states are representational (based on a general framework of what 'representing' consists in) together with the observation that cognitive science tends to use 'representation' when referring to this phenomenon (and the motivations pointed out earlier) underpin my rejection of the arguments by Burge, Sterelny, Cao and others and vindicate the idea that the brain is pervaded with full-blown representations.

This is the answer to the first conceptual difficulty pointed out earlier. Let us move on now to consider the second worry: how can sender-receiver structures be identified in the brain?

4.1.4 *Second difficulty: Sender-Receiver Systems in the brain*

The second question set up at the beginning, how to identify sender-receiver structures in complex neuronal structures, can be divided into two different issues. On the one hand, there is the metaphysical question: are sender-receiver structures instantiated in the brain? The second question is epistemological: how can we pick them out and (even harder) establish their functions? Let us address these questions in order.

4.1.4.1 *Metaphysical Question*

The primary difficulty in finding out instantiations of Third Teleosemantics is not the lack of systems that exemplify a sender-receiver models but their pervasiveness. Indeed, it is sometimes pointed out that there are too many systems in the brain that could qualify as sender-receiver structures. For instance, Cao (2010, p.60) claims:

> The first task is to determine how to identify senders (if any), receivers and signals in the brain. There are several obvious candidates for signals: action potentials, sprays of neurotransmitters, and other non-synaptic molecular messengers (e.g. nitrous oxide). Signals can travel between neurons, or along a single neuron, or between peripheral sensory transducers and primary sensory areas in the brain, or between motor areas and muscles in the body. But boundaries around receivers can be drawn in a almost infinite variety of gerrymandered shapes to include fewer or more constituents.

Cao's argues that in the brain there are too many states and too many candidates for qualifying as sender and receivers in order for the Sender-Receiver model to apply. Certainly, I agree that there are many parts of the brain that can be thought to implement a sender-receiver structure. But I doubt that there is an 'infinite variety of gerrymandered shapes' that can qualify as such (otherwise, neuroscience wouldl be impossible). Indeed, I take it that one of the tasks of neuroscience is to identify these structures, i.e. to group the different parts of the brain into more or less unified units and describe the different patterns of interaction between these structures. In some cases it might be hard to determine significant units, but for the most part cognitive science seeks to adequately describe the different parts in which the brain is divided and its relations. So, of course, the brain is extremely

complex; nevertheless significant systems and states between them can in principle be identified. Why should we think that assuming that there is a great amount of sender-receiver structures instantiated in the brain is problematic? Most people paying attention to actual neuroscientific research think that systems and representations can actually be found out. Let me point at some examples.

Eliasmith (2000, p. 64-5), for instance, identifies representations at different levels: the vehicles of basic representations are single neurons, but he convincingly argues that one can find representations at the level of small groups of neurons (e.g. motion direction, color value,...), which in turn might be part of representations involving larger sets of neurons (e.g. large areas of V1). Similarly, Kandel et al. (2000, p.30), distinguish four signals in each neuron: the input signal (a receptor or synaptic potential), a trigger action (that produces an action potential when a certain spike threshold is reached), the action potential and the output signal (the transmitter release). DeCharms and Zador (2000, p. 624) identify many possible signals, like fire rates of single neurons to firing rates of organized populations of neurons.

I think that, at most, the complexity pointed out by Cao shows that the brain is a semantic engine; there is a complex set of representations across the brain and describing all these systems and representations is an enormous task. But nobody said that doing cognitive science would be easy.[4]

In this thesis, I will only try to identify some sender-receiver structures and hence some representations that are useful for our purposes. I will describe some cases in which one can plausibly find representations at the implementational level, but I will mostly focus on representations at the computational level. Nevertheless, it is important to remark that there might be many ways in which the brain instantiates sender-receiver structures and a high number of states that could be regarded as representations. Some of them may even lie below the level of neurons (i.e. chemical structures). I do not take that to be a problem but one of the consequences of the outstanding complexity of the human brain. In that respect, a huge and very interesting project that this perspective opens (which obviously I will not to try to pursue here) is to investigate all the different ways in which sender-receiver structures are instantiated in different parts of the brain.

Let us move now to the second issue raised by the existence of sender-receiver structures within the brain. Even though we have good reasons for thinking that there are sender-receiver structures in the brain, how can we identify them and their functions?

4 Still, one might worry that there are certain cognitive processes that are hard to accommodate with the sender-receiver model. For instance, one might wonder whether the existence of loops or top-down influences is at odds with the framework set up in Third Teleosemantics.
Now, despite appearances, I think the sender-receiver structure is extremely flexible and can be implemented in many different and surprising ways. Indeed, in this chapter I will explore the idea that senders and receivers can form a chain, in such a way that the system that qualifies as a receiver for a given representational state can be said to be a sender for the next state. Similarly, one could suppose that the receiver of a certain state actually sends a signal back to the previous sender, which becomes a receiver of the new signal. This structure would implement a loop and, at the same time, satisfy the conditions spelled out in Third Teleosemantics.
In that respect, there are many different structures and phenomena that would be interesting to explore with the framework set up in part I. Unfortunately I will only be able to address some of them in detail.

4.1.4.2 *A Methodological Principle*

One of the basic assumptions of this dissertation (shared by the most part of analytic philosophy) is that the method employed by current sciences is about the best we have in order to discover the existence and workings of any feature of the natural world. So the goal of this section is not to reveal a new methodology on how to approach the representational phenomena in neuronal systems; rather, I would like to reflect on how cognitive science actually proceeds and whether this methodology is in fact compatible with the teleosemantic insights defended here.

Two elements seem to be required in order to apply the Sender-Receiver framework suggested here, namely (1) discovering sender-receiver systems and (2) establishing their functions. Concerning the first issue, I already argued that one of the jobs of cognitive science is to discover and classify structures. While this is an herculean task, I doubt that there is any conceptual issue that might conflict with teleosemantics, so I do not think there is anything interesting I can say about it.

But, if the question of finding out sender-receiver structures seems to be hard, research on the etiological function of these systems seems to be almost impossible (see Klein et al., 2002). Without any doubt, one of the key difficulties in carrying out neuroteleosemantics is that, even if in many occasions a certain consumer system for a state can be found out, it is often hard to know its function, and thus, what it needs. Of course, in principle if the brain structure that we identify as a receiver has been selected for (as surely a vast majority are), then it has a function and certain circumstances are required for it to perform its functions in a Normal way. But in many cases, we might not know with great precision what are these needs of the consumer. And, given Third teleosemantics, if we do not know what a consumer system needs, it is impossible to know what a given state represents.[5] How does neuroscience deal with this problem? Is the methodology employed by science incompatible with the teleosemantic approach laid down so far? These are the two main questions that I would like to resolve.

Think first about how neuroscience proceeds. How do neuroscientists know what a particular brain structure represents? Since the origin of modern neuroscience, a common strategy has been the one described by Hubel and Wiesel (1959, p. 574):

> In the central nervous system the visual pathway from retina to striate cortex provides an opportunity to observe and compare single unit responses at several distinct levels. Patterns of light stimuli most effective in influencing units at one level may no longer be the most effective at the next. From differences in responses at successive stages in the pathway one may hope to gain some understanding of the part each stage plays in visual perception.

Generally, one of the central strategies that allows neuroscientists to gain insight into the representational properties of brain structures consists in investigating the cues that most strongly elicit a given state. The stimulus that most intensively activates a given neuronal structure

5 It is worth stressing, though, that the problem is methodological, not ontological. We need to devise a methodological strategy for dealing with the complexities of neuronal structures. That is what I will try to offer in this section

is regarded as the cue that this set of neurons are supposed to gather information about. Thus, neuroscientists usually assume that if a cue causes a high neuronal stimulation, then this is the cue that this brain structure is supposed to track (Eliasmith, 2000; Jacob and Jeannerod, 2003; but see Sheery and Schachter, 1987, p. 439).[6]

As might be obvious at this stage of the dissertation, in the context of our discussion the appeal to 'cues that elicit the representation' is too imprecise. Surely, neuroscientists do not think that a brain structure represents *any* cues that elicit a representation, not even most of them. This strategy would not be very illuminating, since probably almost anything can cause activation of a certain neuronal structure *in some circumstances.* And, obviously, in order to solve this problem, we cannot appeal yet to the stimulus that the neuron has the function to detect (in the sense of Etiological Function), since the methodological problem originated precisely because in many cases we ignore the details of the evolution of a given structure.

We get, then, to three desiderata our account must try to comply with: (1) capture the way neuroscientists proceed, (2) fulfill 1 without contradicting what I defended in the first part of the dissertation and, as the same time, (3) solving the problem of picking out the right kind of entity a neuronal structure is supposed to represent.

This is a complex challenge that, I think, can be met. In chapter 1 I presented an account that I suggested could be used (with some slight but important modifications), in order to describe the way neuroscientists investigate the brain. In particular, I described Rupert's approach, which claims:

> Relative Indication 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 than are members of any other natural kind.

Notice that Relative Indication partially captures the leading intuition of neuroscientists: the represented cues are the cues that are more efficient in causing a given mental state. However, three features of Rupert's view are inadequate for the task at hand. First of all, Relative Indication is supposed to be a (semantic or metasemantic) theory of content. I have already shown that this account cannot be right as a semantic or metasemantic theory of content (see 1.2.3.3). Nevertheless, it can still do some job; given that we are just trying to devise a methodological strategy for addressing sender-receiver systems in the brain, something like Rupert's theory can be used as a methodological principle.

The second flaw of Relative Indication is that it is restricted to members of natural kinds, but this condition makes no sense in the present context, since brain structures are usually said to represent properties like *such and such wavelengths impinging the retina, being red, being cylindrical*, and so on.

Thirdly, Relative Indication faces the difficulty pointed out in chapter 1: we could device artificial and gerrymandered objects ('superstimuli') that activate a brain state more strongly than any other feature. In order to solve this worry, we should consider only properties that plausibly were around the organism when these brain structures

6 This is true, as least, with respect to research on perceptual systems. We will see that higher-order representations might be studied differently (see 6.2.2.1).

evolved. In other words, we should try to approximate the Normal conditions as much as possible. Even if we ignore the exact function of a given system (see above), we might nevertheless have a rough idea of what kind of cues where around the organism when and where it evolved.

Hence, I suggest the following modification of Rupert's account:

RELATIVE INDICATION* As a working hypothesis, assume that R has as its extension the members of Q if and only if:

1. It can plausibly be assumed that members of Q were present in Normal circumstances.
2. Members of Q are more efficient in their causing of R than are members of any other kind.

I think that this principle roughly captures the strategy used by cognitive scientists (Eliasmith, 2000; Jacob and Jeannerod, 2003; DeCharms and Zador, 2000). Furthermore, it provides an extremely useful methodological principle for developing neurosemantics.

Now, one might worry about the justification for such a methodological principle. I said that this methodology employed by neuroscience yields the right results in the overwhelming set of cases. But, why should we think that in general the cue that the most strongly elicits a given brain structure in accordance with RELATIVE INDICATION* is indeed the represented feature according to THIRD TELEOSEMANTICS? There are two broad motivations for that strategy.

First of all, as a matter of fact, in many cases the result provided by RELATIVE INDICATION* coincides with the content that THIRD TELEOSEMANTICS yields (some examples will be considered below). Thus, RELATIVE INDICATION* is vindicated (again, as a methodological principle) by many cases where it turns out that the represented feature (in accordance with THIRD TELEOSEMANTICS) is precisely the state that produces a high activation of the neuronal structure. In the case of the human visual system, for instance, it is very likely that the content of the many neuronal states roughly corresponds with the cues the system is able to discriminate, because as we will see these are the properties that are relevant for the consumer system to perform its functions in a Normal way. Hence, the strategy consists in provisionally accepting as a working hypothesis that in sophisticated representational systems the cues that the system is able to discriminate correspond with the properties that the consumer-system needs in order to perform its functions Normally.

Secondly, one could try to develop a more general argument for motivating this methodology along the following lines. Paramecia identify oxygen-free areas by sensing geomagnetic north, but the connection between geomagnetic north and oxygen-free areas is very distal. Frogs identify good flies by sensing black moving things, and the connection between *being a fly* and *being a black moving thing* is quite close. In turn, the human perceptual system identifies (say) flies by sensing their being small, having certain texture, color, shape of wings,.. So, as a general rule of thumb, one might expect that the more sophisticated a cognitive system is, the closer is the connection between the cues that elicit the representation and the represented state of affairs. In other words, if frogs had evolved a more sophisticated cognitive system, they would very probably have evolved a mechanism that would be activated by more specific properties of flies. Of course, this is an empirical

claim, but as far as one finds this line of reasoning compelling, it lends support to the methodological procedure suggested here. If it is true that the more sophisticated a representational system is, the closer is the state of affairs that a state discriminates from the state of affairs that a state represents, then we can expect that in highly sophisticated representational systems like the human brain, both things will usually come together. So, in this complex systems, Relative Indication and Third Teleosemantics will often yield the same result.

Let me stress that Relative Indication* just provides a methodological strategy. That means that if the working hypothesis that results from Relative Indication* clashes with what we know about the functions of the systems, the latter is obviously the determining feature. If the state of affairs that most strongly elicits a given activation in a brain structure (in the sense of Relative Indication*) does not seem to be the state of affairs that the consumer system has Normally required in order to perform its functions, then *the needs of the consumer system always prevail*. In other words, in case of disagreement, the consumer system utterly determines the content of the representation. This is just a different way of saying that in this section I am only setting up a methodological principle; content is determined by the conditions set up in Third Teleosemantics. This is why nothing I say here contradicts the view carefully set up in the first part of the dissertation.

Interestingly, the very same idea I defend was hinted at by Bechtel (2000, p. 338, emphasis added):

> Althought I am arguing that the focus in constructing a state or event as a representation is on the consumer of the representation, neuroscientists typically begin by trying to correlate neural activity with external processes that they might represent (...). This is, indeed, quite sensible and does not undercut my claim. An extremely useful first step in determining what the system takes a state or event to represent is to ascertain what information a state or event might carry. Then one asks the question of how the system was designed to use the information.

It is time to take stock. In order to address the problem of identifying systems and functions in perception and cognition, I suggest the following methodological strategy:

Procedure

1. First, consider how the producer system generates a representation. In particular, find out which stimulus Q most strongly elicits a given mental state, in accordance with Relative Indication*. As a first hypothesis, suppose that this system represents stimulus Q.
2. Secondly, find out whether it is plausible to hold that this state of affairs is what the consumer system needs in order to perform its own functions in a Normal way, as stated in Third Teleosemantics. The latter is what really determines content, but since the needs of the consumer system are often hard to assess, the best working hypothesis we have when addressing complex systems is that a particular brain structure

represents whatever it is sensitive to. Once we know what a system most strongly reacts to, we should consider whether it is reasonable to hold that this state of affairs is what is Normally needed for the consumer to perform its functions in a Normal way.

3. Finally, if we have good reasons for thinking that this state is not what the consumer system needs, then we will have to reconsider the content of the representation in light of the needs of the consumer-system. The motto is the following: *in case of disagreement, the needs of the consumer system prevail.* That means that, in some situations, what a producer system is sensitive to might not qualify as the represented state of affairs because the needs of the consumer system are different. Nonetheless, I pointed out some reasons for thinking that the methodological strategy will probably be useful because, very often, the state that satisfies Relative Indication* will be the state that satisfies Third Teleosemantics.

With Procedure we solve three problems at the same time: on the one hand, current practice in neuroscience can be understood and accommodated within the teleosemantic paradigm I suggested. Secondly, this proposal will provide a set of working hypotheses about the representational properties of neuronal structures, which would have been very hard to obtain by exclusively relying on Third Teleosemantics. Finally, by appealing to (plausible) Normal conditions we exclude most artificial stimuli like the superstimuli described in chapter 1.

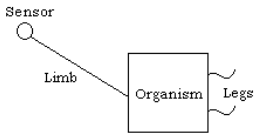
Furthermore, notice that this strategy does not amount to adopting Neander's 'Low Church Teleosemantics' (see 3.3.1), which is the strategy that most teleosemanticists working in neuroscience have assumed (Neander, 2006; Jacob and Jeannerod, 2003; Eliasmith, 2000). Paying attention to the production of the representation just is a methodological strategy for solving the problem of finding out plausible hypotheses about evolution and consumer systems in cognitively sophisticated organisms. As I argued at length, I think that consumer-based teleosemantics is the only proposal that can deal with the four difficulties put forward in chapter 1, 1.2.2.1. Nonetheless, I think that relative indication* provides a useful methodological strategy, which is indeed extensively used in cognitive science (Jacob and Jeannerod, 2003; Neander, 2006). If certain tension between the two methods arises (we will see that this will not happen very often), we will obviously side with the perspective that takes into consideration the needs of the consumer system. This, of course, was one of the conclusions of the first part of the dissertation.

Having addressed the preliminary questions 1 and 2 and set up a useful methodological strategy (Procedure), let us consider in detail some illustrative cases in which the neuroteleosemantic proposal applies.

4.2 perceptual systems

In the previous sections I have directly addressed conceptual issues related to the first two concerns raised at the onset. In this section, I

Figure 1: Creature designed by Mandik (2003).



would like to fill in a neuroteleosemantic account by describing in some detail certain empirical cases. In what follows, I will put forward in detail three examples of structures that implement a sender-receiver model, and argue that everything we have said so far seems to be true of them. That will show that the framework outlined in the first part of the dissertation is actually exemplified in some real cases. The examples are presented in order of increasing complexity.

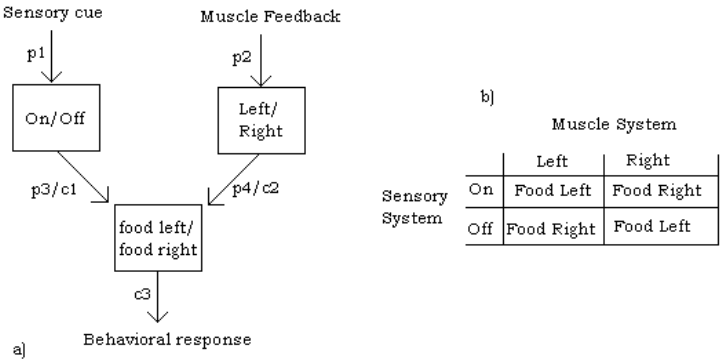
4.2.1 *Computer models*

According to our definition, a sender-receiver structure is instantiated whenever two mechanisms have been shaped by natural selection in such a way that the function of the producer is to indicate the consumer that certain circumstances obtain, in order for the consumer to perform its functions in a Normal way. But, so far, we have only discussed cases where the consumer system was supposed to behave in a certain way towards certain environmental feature (predators, food, oxygen-free area,...), and we have not considered cases where the consumer's job is to produce a further internal state of the organism. How can the sender-receiver framework be applied here?

First of all, it is worth mentioning that some abstract models in Signal Theory include signaling systems in chain. For instance, Skyrms (2010) discusses representational systems mediated by a translator, that is, a mechanism that receives certain signs as inputs and yields other signs as outputs ('translated', so to speak). Similarly, he discusses complex signaling structures such as informational rings or 'star figures' (Skyrms, 2010, p. 149- 177). For our purposes, the important idea is that a chain of representational systems is not a conceptual innovation in the field. Here, I will focus my attention on a computer model which illustrates many of the issues I have been presenting.

A simple model shows how the sender-receiver framework can be instantiated in a system composed of different sub-systems. Mandik (2003) has designed a series of computer models of neural networks where he tested different theories about the origin and evolution of simple representational systems. In one of his simulations, he employed a creature with a single sensor mounted on a long limb that was used as an oscillating scanner (Mandik, 2003, p.118). He designed a creature that was endowed with (1) an antenna, which had a sensor at the end that could be in two states (on and off) and (2) a feedback mechanism that indicated whether the limb was on the right (state being on) or on the left (state being off) (see figure 1). The creature also possessed a pair of legs, such that it could propel himself through the waterish medium.

Figure 2: Systems and representations of Mandik's AI model.



As Mandik (2003, p. 118-9) describes the creature, the sensory activity was supposed to encode proximity information (it was supposed to fire when certain cue was present), while the feedback coming from the limb provided information about the limb's position. By computing the information provided by the sensory activity and the muscle feedback, the organism could move in the direction of the food. For instance, if there was sensory activity and the muscle was bending to the right, then the food was to the right so the organism moved in that direction. In contrast, if there was sensory activity while the muscle was bending to the the left, then the food was to the left, so it moved accordingly. If there was no sensory activity while it was bending to the right, it moved to the left in order to find food and if there was no sensory activity while the muscle was bending to the left, then it moved to the right (See figure 2.B).

Now, there seem to be three representational systems involved in this model: the sensory system, the muscle system and the motor system. Each system instantiates a sender-receiver model, and hence produces a representation. As the model was designed, the sensory sender-receiver structure and the muscle sender-receiver structure existed because they allowed the motor system to compute the location of food. That is, the function[7] of the consumer systems in the sensory and the muscle system was to produce a further representational state that allowed the motor system to act appropriately. Once the functions of the sensory and muscle consumers are established, then we know what these systems need in order to perform their tasks appropriately, and (following the consumer-based teleosemantics defended here) we also ascertain what are the functions of the producers. In the sensory system, the function of the producer (p1) is to generate a state that corresponds with a strong intensity of certain cue. On the other hand, the function of the producer in the muscle feedback mechanism (p2) is to produce a state that corresponds with the certain position of the limb. Only when the sensory representation corresponds with a certain cue being

7 Of course, these states do not have functions in the sense of Etiological function, since they are computer models. They have functions (in some sense to be elucidated) in virtue of being designed with a certain purpose. Since my goal in this section is only to illustrate how complex systems can instantiate multiple sender-receiver structures, this point is irrelevant. Later on we will see real examples of similar systems that are endowed with functions in the sense of etiological function.

present and the muscle representation correlates with the limb being in a position can the motor system rightly compute the location of food and, hence, all these states lead to a successful behavior.

In turn, the function of the consumer of the motor system (c3) is to bring the organism to the food source. So it needs to represent the location of food. The inputs by means of which it generates a representation of the location of food are the representation issued from the sensory system (which indicates the strength of the cue) and the representation produced in the limb system.[8]

Crucially, notice that the content of the representation in the motor system is not the same as the sum (or some other sort of composition) of the contents of the sensory and muscle representations. While the content of the sensory representation is something like *there is an intense cue* and the content of the muscle representation is *limb is at position L,* the content of the motor representation is *there is food at location L.* The content of these states is different because the needs of their respective consumer systems are also different.

Interestingly enough, there are certain mechanisms in living organisms that resemble very much Mandik's model. Cockroaches and crickets, for instance, have two short appendages that extend from the rear of their abdomen called 'cerci'. Each cercus has a set of 'slender filiform sensory hairs' Comer and Lung, 2004, p. 314), which are sensitive to air movements. Each hair is associated with an efferent neuron, such that when a particular hair senses air moving at a certain velocity (which usually enough corresponds with the presence of a predator), cockroaches respond with an evasive behavior in a certain direction.

This example highlights several other important issues. On the one hand, it shows how sender-receiver structures can be found within simple cognitive systems. On the other, it illustrates how sender-receiver structures can form a chain, such that the function of one consumer is to produce a representation of a cue needed by the next sender-receiver structure (that was one of the problems indicated in 3 above). Thus, sender-receiver systems need not form a close loop in the way Cao (2012) envisaged. Consumers need not always be some kind of motor system, which elicits certain behavioral response to a cue that has caused the representation. The connection between external input and behavioral output need not be so tight. Several representational systems might be connected in a row. Furthermore, different systems in the chain might be representing different features. That suggests that there might be some room for neuroteleosemantics of more complex cognitive systems.

4.2.2 *Toad cognition*

Now it is time to consider how the teleosemantic framework outlined here can be applied to a relatively sophisticated cognitive system. In particular, before moving to human cognition, it might be useful to

8 In figure 2, the consumer of the sensory system (c1) is the same as the producer of the motor system (p3), and the consumer of the muscle mechanism (c2) is the same as the producer of the motor system (p4). In general, the distinction between a consumer system of one system and the producer system of a different system might no be clearly cut-off. For instance, if a box is connected by a wire to a different box (which can be on or off), it seems that the linking wire qualifies as part of the consumer system of the first box and as the producer system of the second box. Hence, this is a feature of Mandik's model that can also be found in other structures. In other situations, however, the two systems can physically come apart.

consider the perceptual system of toads (*Bufo bufo*), which is simple enough for being tractable and, at the same time, complex enough for providing a first approximation to the human mind.

There is a striking feature of toad cognition that makes it an interesting case to discuss at this point of our research. Even if some of the toad's cognitive abilities are slightly flexible due to the fact that (1) they can be habituated by repeated exposure to a certain stimulus and (2) the strength of certain behavioral responses depend on their motivational states, the mechanisms I will focus on are indeed fairly automatic. So even though toads have some capacity of learning (for instance, by associative conditioning; Ewert, 2004, p. 144), given a stimulus of a certain kind, the behavioral response is triggered with a high probability. At many stages, once a given sequence of processing has begun, it cannot be modified in light of further information. Using Sterelny's terminology, their states qualify as *coupled representations*. We will leave the question of decoupled representations (states that are not designed to produce any specific behavior) for the case of the human perceptual system.

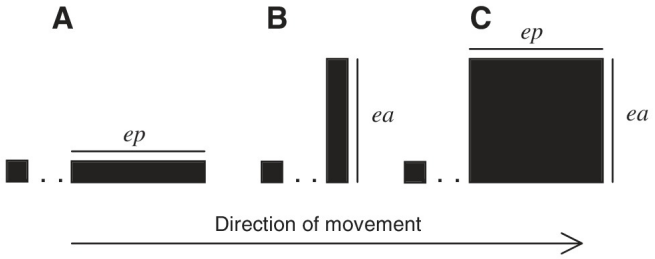
Despite the fact that the toad's perceptual mechanisms are much simpler than the human's, it is important to note that toads already have some of the sophistications of human cognition: a complex visual system, composed of a set of ganglion cells, optic nerve, and so on, as well as a certain degree of computation. Furthermore, even within the toad's visual system, there are several dissociated pathways (Ewert, 2004, p. 140). If I am able to show that teleosemantics applies to this cognitive system, we will move an important step towards a neuroteleosemantic account of human perceptual systems.

Let me start by providing a description of some relevant aspects of toads (*Bufo Bufo*). Toads may have different diets, depending on the species and the size of the individual, but they generally eat a variety of things, such a beetles, bugs, slugs, millipedes flies and earthworms. All these organisms have one feature in common- they all move in the direction of their longer body axis (Ewert, 2004). Nonetheless, toads can also represent other features that are relevant for their survival. Apparently, the toad's visually induced behavioral responses roughly discriminate between three different kinds of stimuli, what we could (tentatively) describe as 'prey', 'predator' and 'mate' (Ewert, 1999, p. 172). The standard response to each of these categories is to catch them, avoid them and approach them, respectively. Here I am going to focus on prey-detection, but I will show how the approach presented here can also accommodate states that generate predator-adequate responses.

There are many responses toads perform in light of a predator (sidestepping, ducking, crawling...) but the behavioral responses to prey-stimulus are usually more limited, and researchers classify them under four basic categories: orienting towards the stimulus (o), stalking at approaching (a), viewing the prey from the front (v), snapping at it (s). These behaviors are usually performed in a sequence, where some actions might be repeated until the toad considers he has reached an adequate position. For instance, a possible sequence of those behaviors is the following: o-o-a-a–o-o-v-a-a-s (Neander, 2006).

But, how do toads recognize the presence of prey? In a famous series of studies, Jorg-Peter Ewert and his colleagues used a set of artificially designed stimuli (including cutouts) with three distinct configurations, that were intended to examine what kind of configuration elicits the

Figure 3: Configurations employed in Ewert's experiments.



stronger behavioral response in toads. The three configurations consisted of (A) rectangles of constant width and varying lengths moving in a direction parallel to their longest axis, called 'worms' (B) rectangles with constant width and varying length moved in a direction perpendicular to the longest axis, called 'anti-worms', and (C) squares of different sizes moving in the same direction as worms and anti-worms (see figure 3)

The experiments showed that toads react much more strongly to worms than to anti-worms or squares (Ewert et al. 1999; Ewert, 2004). In particular, it was observed that toads react to a configuration that is composed of a certain shape moving in a certain way. Neuroethologists emphasize that neither of the two stimuli alone (either shape or movement) is able to elicit a strong response.

This was the behavioral evidence that was required for investigating the neural substrate of such behavior. Notice that the methodology used by Ewert and colleagues matches perfectly well with the strategy set up in Relative Indication*. In order to come up with a certain hypothesis about the representational content of certain brain states, they devised a set of different stimuli and did some experiments in order to see which of them more strongly elicit neuronal activity. And, of course, the kind of stimuli they used were supposed to approximate the kind of cues that frogs usually find in the wild.

Let us move now to the neurological basis that underlies this capacity of toads.

4.2.2.1 *Neural basis for prey-detection*

The processing of visual information begins in the retina. As in many other organisms, the toad's retina is composed of rods and cones, which transduce light into neural firings. Rods and cones are two different kinds of cells that fire in response to different stimuli. They in turn activate ganglion cells, that compose the optic nerve and extend to the mid-brain structures. All ganglion cells have receptive fields composed of two concentric fields, an excitatory inner circle and an inhibitory outer circle. That means that they fire more strongly when there is a contrast between a stimulus in the center and the outer circle. Each cell responds best when the entire center is stimulated and none of the surround is.

However, not all cells have the same kind of receptive field or fire with the same intensity. There are four different sorts of ganglion cells, R1, R2, R3 and R4, which differ with respect to size of the excitatory

receptive field, the strength of their inhibitory receptive fields, the velocity of the stimulus and the contrast between excitatory and inhibitory stimuli (Ewert, 1997,2004). The details of the differences between R1, R2, R3 and R4 need not concern us.

An important feature of these four types of ganglion cell is that it seems that the activity of neither of them straightforwardly maps onto the behavioral responses suggested in the previous section (i.e. orienting towards the stimulus, stalking at approaching, viewing the prey from the front and snapping at it). The fact that there is no simple mapping between the activity of these different cells in early vision and the toad's behavioral responses suggests to neuroethologists that the discrimination of prey, predators and others requires further processing.

The ganglion cells primary extend to the optic tectum, a mid-brain structure and also to the thalamic pretectum, in addition to other neuronal structures (Ewert, 1997). As in the human visual system, there is crossing-over between the ganglion cells coming from the left and the right eye. Furthermore, neighborhood relations are preserved, which means that ganglion cells that have nearby receptive fields (or rather: largely overlapping receptive fields) project onto nearby parts of certain visual layers. So these layers in the visual system form a retinotopic representation.

Now, after extensive research in this complex visual system, neuroethologists have discovered a set of neurons in the optic tectum called T5.2 that qualify as the best candidate for correlating with worm-like stimulus. If one compares the activation pattern of these cells in response to worm-like, anti-worm and square stimulus, one can find a surprising match between this activity and the behavioral responses discovered by Ewert at al.:

> Tectal T5.2 neurons differentiate between [worm-like] and [anti-worm figures]. Their responses to changes in [worm-like] and [anti-worm figures] resemble the toad's prey-catching activity (...). In this respect, T5.2 neurons are prey selective, which could be explained by excitatory input from T5.1 and inhibitory in put from TH3 cells. (Ewert, 2004, p. 136)

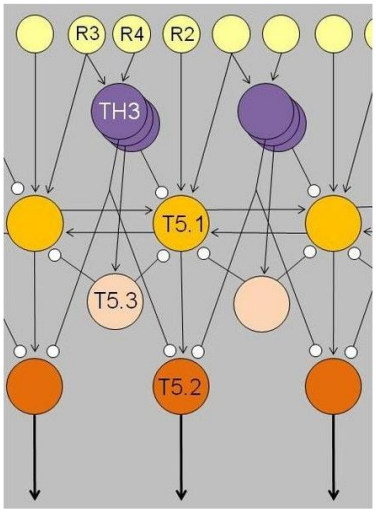
Furthermore, their axons project down to the spinal motor systems which harbor the motor neurons of the tongue muscles (Ewert, 2004, p. 136). This neurophysiological evidence fits very nicely with the idea that T5.2 are responsible for prey-detection.

On the other hand, T5.4 neurons display a sensitivity to large and compact objects, which elicit avoidance and escaping behavior. In fact, for every behavior of the toad, one can find a set of neuronal states that elicits the behavior. Ewert (2004) calls this view 'the hypothesis of sensorimotor codes', which is a different way of talking about Tinbergen's (1960) 'releasing mechanisms'. Figure 4 (extracted from Ewert, 2004) illustrates the structure of the toad's early vision.

This picture includes other kinds of neurons, such as inhibitory neurons (TH3), excitatory neurons (T5.1) and several connections that make the mechanism a bit more complex. I would like to focus, however, on T5.2 neurons. With respect to them, Ewert claims:

> In this language, the sensorimotor code of a command-releasing system embodies a perceptual schema that exists for only one purpose: *to determine the conditions for activation*

Figure 4: Neuronal connections in the toad's early visual system.



> *of specific motor pattern* generator embodying a motor schema. (Ewert, 2004, p. 149, emphasis added).

In this quote, Ewert seems to be in complete agreement with a neuroteleosemantic account of these mechanisms: these states represent the condition that is required for the motor response to be successful. Let us present in more detail how teleosemantics can explain this case.

4.2.2.2 *Neuroteleosemantics in Toad's Visual System*

Now it is time to apply the teleosemantic framework developed so far to the visuomotor system in toads.[9]

As we saw, the first element we need in order to apply Third Teleosemantics is to identify several places where a Sender-Receiver structure is instantiated. That is, we must find structures that have been selected for performing certain tasks defined in third Sender-Receiver. Let us remember the definition we gave at 3.2.5:

> Third Sender-Receiver Any two systems P and C configure a sender-receiver structure if, and only if:
>
> 1. P and C have functions in accordance with Etiological Function
> 2. P and C have coevolved in such a way that a Normal condition for the proper performance of each system is the presence and proper functioning of the other.
> 3. P has two functions:
>
>   a) The non-relational function of helping C to perform its functions.

9 Let me stress that, at that point, I am probably partially disagreeing with Millikan. At certain places she has claimed that states in early vision do not qualify as representations: "Of course, retinal patterns themselves are not perceptions or intentional signs of anything, certainly not just as such" (Millikan, 2004, p.55). Nevertheless, consider this other quote: "For example, the edge detector cells in early vision represent edges, not light intensity gradients across the retina. Their function is appropriately to guide internal acts of identification of contours and shapes, given the presence of certain edges." (Millikan, 2004, p.83).

b) The relational function to produce a set of states R, which are supposed to map onto a set of states S in accordance with a certain mapping function *f*.

4. The function of C is to produce an effect E. The most proximal and most comprehensive Normal explanation for C's performance of E involves members of S.

Let us take a particular step in the functioning of the toad's perceptual mechanism and show how this schema applies to these processes.

RETINAL GANGLION CELLS For instance, let us concentrate on the ganglion cells R2, which are activated by the photoreceptor cells in the toad's retina. As a first approximation, these cells seem to be the producer systems that generate representations, the representations seem to be the activity produced by these cells (probably a certain firing rate) and the consumer seems to be the set of cells that receive this activity (T5.1, as suggested in Figure 4).

Let us go through the conditions set up in THIRD SENDER-RECEIVER and show why they indeed satisfy all conditions for instantiating the sender-receiver model:

(Condition 1) First of all, it is undeniable that R2 have functions in the sense of ETIOLOGICAL FUNCTIONS. Scientists surely assume that these mechanisms have been selected for that particular task. Arguably, ganglion cells exist because sometimes they produce certain activity (basically, an action potential at a certain rate) that allow the next process downstream to perform other tasks.

(Condition 2) Secondly, it is also obvious that all these cells are cooperating mechanisms, in the sense that they have coevolved in such a way that (in Normal conditions) they function properly only when the other mechanism obtains.

(Condition 3a) 3a seems also to be easily satisfied: the function of these cells is to help the next group of cells to perform its functions.

(Condition 3b) Apparently, one of the functions of ganglion cells R2 is to fire at a certain rate when a stimulus with certain features is presented. Notice that for every kind of retina cell, there is a different mapping function from action potentials to stimuli. In other words, the same neuronal activity in R1 and R2 represent different states, since these states have been selected for correlating with certain states of affairs in accordance with different mapping functions (that will be discussed in more detail below). For instance, high activity in R1 corresponds to an object of 3-4 mm, while high activity in R2 represents an object of 5-10 mm (Ewert, 1997, p. 333). Evolution has designed different cells that are supposed to correspond with different features. The key issue is that firing rates map onto the presence of certain stimuli.

(Condition 4) Finally, the fact that these stimuli in fact obtain seems to be the Normal condition for the next set of cells to perform their function in a Normal way. Suppose R2 fires at

> a high rate but there is no object between 5-10 mm. Then, cells T5.1, which according to the diagram receive activity from R2 and are supposed to produce a representation of an object with certain features moving in a certain way, will probably produce a false representation. The fact that the entity represented by R2 obtains is the Normal condition for the next mechanism upstream to produce its own true representation. Generally, a precondition for the next step in the computational process to perform its own function in a Normal way is that the activity in early stages of vision corresponds with the presence of certain features. So condition 4 also seems to obtain.

Thus, it seems that we have localized a sender-receiver structure; the function of the ganglion cell structure R2 (or, more precisely, of the structure which encompasses the dendrites, axon and the body cell) is to generate a state (firing at a certain rate) that is supposed to correlate with certain environmental feature. The fact that the representation actually correlates with the occurrence of a certain state of affairs explains why the next level of computation can produce a representation of another state and, hence, why it can fulfill its function in a Normal way. More generally, if the firing of every ganglion cell counts as a representation, the set of ganglion cells R1-R4 produce a complex representational state, where every cell indicates the presence of a certain cue at a certain location of its receptive field.[10]

THE PICTURE The key insight that will be developed here is that the same teleosemantic analysis can be provided for every computational step in the perceptual system, so that a large amount of Sender-Receiver structures can be identified. The task of every set of neurons (R1, R4, T5.1,...) is to produce a representational state that corresponds with a certain environmental feature, since this is a condition that the next assemblage of neurons requires in order to perform its own function in a Normal way (which probably consists in computing the occurrence of a different feature). In this way, we can interpret cognitive computation as a reiterative instantiation of sender-receiver structures, where every set of neurons needs to produce a representation whose content has to be true for the next level to produce its own true representation.

The idea that brain representations originate within a sort of sender-receiver systems instantiated in the brain has been hinted at by some people in different fields. For instance, neuroethologists Staaden and colleagues (2004, p. 335) claimed:

> An animal's perceptual representations of the world is the product of sensory systems that have been shaped by natural selection. The evolution of the signal component of the signaler-receiver relationship inherent in the sensory systems that feeds the animal's perceptions has been described as an "economical process" that is unlikely to have arisen in

10 A similar sort of complex structure is found in many other transducers, e.g. the skin thermoreceptors, which enable us to detect temperature (Akins, 1996). There are also different kinds of thermoreceptors (cold vs. warm receptors), which fire at different rates and are distributed differently in the organism. The activity of all thermoreceptors forms a sort of map of the temperature around the organism. If teleosemantics can account for the particular case of vision, it is to be expected that it will also work for many other sensory systems (cfr. Akins, 1996).

> isolation but, rather, which has taken shape in terms of the background stimuli against which the signals exist.

A similar idea was suggested (but not developed) by Bechtel (1998), who adopted a sort of consumer-based teleosemantics. He argued that current neuroscience actually assumes that certain brain events are representations in virtue of being consumed in a certain way by the mechanisms situated upstream in the computational hierarchy:

> Investigators would have had little interest in figuring out the information relation between brain activity and distal stimuli unless they assumed the brain was using the information in the processing. To determine how the brain actually consumes this information requires developing processing models which show how activations in later areas in visual pathways utilize information encoded in earlier stages of processing. (...). By showing that processes at each stage respond to those upstream to arrive at their characteristic response, they show that upstream processes are representations. (...) Much current work, especially in computational neuroscience, is devoted to developing processing models showing how later visual areas can generate their representations from what is represented in earlier visual areas. (Bechtel, 1998, p. 340-1)

Crucially, notice that so far we have been following the methodology I set up in PROCEDURE (see 4.1.4.2). First, we agree with neuroethologists in assuming that the represented property is the one that more strongly elicits neuronal activity. This is supposed to yield a first working hypothesis. Secondly, we check whether this result provides a plausible condition required by the needs of the consumer system. When both results fit each other (as in the case of cells in early vision), we can tentatively conclude that we have identified a sender-receiver structure and that the content of this state in question is such and such.

Thus, every significant set of neurons in the visual system that takes part in vision computes a different feature configuration (motion at location l_1, a certain edge at l_2,...). At least in the toad's case, this pattern is iterated until we reach the 'releasing mechanism'. When we find the mechanism that puts in motion the behavioral response, the analysis becomes significantly different in an important respect.

RELEASING MECHANISM Consider T5.2, which Ewert and others identify as the toad's releasing mechanism. We saw that this structure leads directly to the motor system. That is, the consumer system for that representation consists in certain neuronal structures that automatically produce a behavioral response to the stimulus (which might also depend on other inputs: motivational state, season,...). In this case, the environmental feature that the consumer system needs in order to act appropriately and Normally is not the presence of any perceptual feature -the motor system is not supposed to create a more complex or accurate representation of colors, figures or movement. What the motor system needs in order for its activity to be successful is the presence of (roughly) a *small insect*. Using the terminology we defined earlier, the most proximal and most comprehensive Normal explanation of how the consumer system of T5.2 performs its function must mention the presence of prey, rather than the presence of a certain

black thing moving around (see the discussion on the indeterminacy problems, in 2.3.3.1). So, as a first approximation, T5.2 has the following representational content *there is a small insect at location l.*

Again, notice that this result follows from the strategy described in Procedure. This time, however, we get to condition 3 of Procedure: in the case of T5.2 neurons, the feature that most strongly elicits activity (the one that satisfies Relative Indication*) does not seem to correspond with the feature that the consumer system needs. The consumer system (in that case, the motor system) plausibly needs small insects (and not black moving things). And since the needs of the consumer systems always prevail, the content is something like *small insect.*

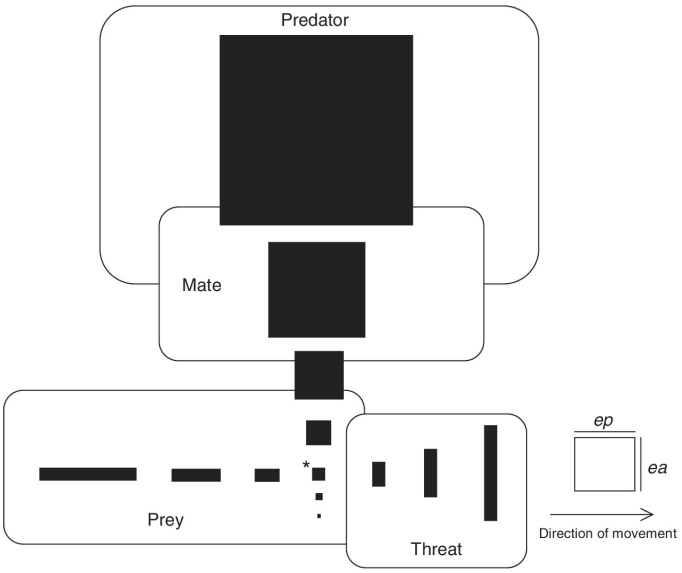
Let me repeat the reasoning underlying the content attribution to T5.2, because it is very important. First of all (following step 1 in Procedure), we considered the stimulus that more strongly elicits T5.2, namely a worm-like figure moving in the direction of its axis. However, in contrast to the previous analysis of representations in earlier stages of the visual system like activations in retina cells, when we focus on the consumer of activity in T5.2 (which seems to be the motor system), this system does not seem to require a representation of a worm-like figure; the motor system works properly only when there is a prey around. Only if there is a small insect around can the catching succeed (see 2.3.3.1). So this is an example where the the working hypothesis that derives from Relative Indication* (that T5.2 firing represents *there is a worm-like stimuli moving along its axis)* is in contradiction with the plausible needs of the consumer system (which requires something like the presence of a small insect). And of course, following condition 3 of Procedure, if the analysis developed in part 1 is right, what really determines content are the needs of the consumer system. So activation of T5.2 means *there is a small insect at location l.*

The result then is the following: there is a whole range of states in early visual processing that represent the presence of certain features: blackness, edges, moving stimuli,.. but, at some point, a particular neuronal structure represents something far more distal: the presence of the insect. This is so because in this case the needs of the consumer system are plausibly not the computation of a further environmental property, but the production of a certain behavior. I think this is a nice way of unifying the lessons of neuroscience with the teleosemantic perspective offered so far.[11]

But, one might wonder, why does T5.2 represent *small insect* rather than *beattle, fly*, or a conjunction of both? First, remember that, in this case, content is determined by the set of properties that must figure in the most proximal and most comprehensive Normal explanation of how the consumer system has historically performed its function in a Normal way. We saw that toads generally eat a variety of things: beetles, bugs, slugs, millipedes flies, earthworms, etc.. and all of them have contributed to the fact that this sender-receiver system (and this particular mental state) exists. But citing all these different kinds of prey would be too detailed a description. The least detailed Normal explanation will not mention the fact that what a toad was preying was a fly or a beetle, in the same sense in which the most proximal Normal

11 This idea constitutes a reply to Akins' (1996, p. 365) challenge against naturalistic theories of perceptual content: a teleosemantic theory can account for the fact that certain perceptual mechanisms represent a distal property by means of representing a proximal feature. Similarly, it could be argued that object-centered representations can be produced through the activity of subject-centered representations and vice versa (see below).

Figure 5: Toads interpret each configuration as indicating the presence of a different kind of entity.



explanation of the selection of the chameleon's pigment-rearranging device does not mention the fact that the skin was green or brown. T5.2 represents a set of features that all these organisms share, whose most comprehensive expression is something like *there is a small insect at l.*

Finally, notice that the same analysis I have carried out concerning the representational content of T5.2, also applies to many other states. As we saw earlier, the toad is endowed with different neurons that are attuned to generate the right reaction when confronted with predators, prey, or mates.

Consider figure 5, which can be found in Ewert (2004 p. 156). For each kind of relevant feature (prey, mate, predator, threat) there is a set of neurons specialized in producing the right behavior. While all these mechanisms roughly use the same input information (size, shape, movement,...) each configuration of features elicits a different behavior that is adequate for a different state of affairs. This is why, even if the sensory information of the frog is limited, it nevertheless can represent *prey* or *mate*.

This is all I wanted to say about toad cognition. This discussion has helped us to show how teleosemantics can be applied to real cognitive phenomena. Let us move now to a much more complex system: human perception.

4.2.3 *Human Perceptual Systems*

The picture that emerges from the previous considerations is that the perceptual system is composed of many sub-systems, and each of these different sub-systems plays a different computational role, approximately following Marr's (1982) computational model, which is still standardly used in cognitive science (Frisby and Stone, 2010). Basically, what visual perception does is to receive all the different information that comes from various inputs (fundamentally, the light impinging the

retina, but also some other information like proprioceptive information about eye location) and to process this rich and varied information so as to get a (more or less) coherent perceptual state that represents a certain complex state of affairs. The general strategy for understanding perception within teleosemantics is the following: each identifiable sub-system in the complex processing instantiates a sender-receiver structure, and hence, according to the teleosemantic model I have been defending, it possesses contentful representations. What we usually call the 'perceptual experience' is but a particular subset of this complex amount of representational states.

Certainly, in order to provide a full neuroteleosemantic account of perception, I would have to describe all the different parts into which the perceptual system is composed, analyze why and how they instantiate a sender-receiver model and show that they can generate a complex representation (perceptual state). Needless to say, providing the full story lays beyond the scope of this thesis.[12]

Nevertheless, I have already shown how a neuroteleosemantic approach can be developed in a real cognitive system such as the toad's perceptual system. So what I need to so is to show how this account can be applied to *human* cognition. I will do that by sketching the main steps of the creation of a visual percept and focusing on those aspects that distinguish the toad's perceptual system from the human's. Afterwards, I will consider some features of human cognition that may pose certain difficulties to my neuroteleosemantic account. Finally, I will defend the view of perception I offer from some objections that can be found within the philosophy of perception.

4.2.3.1 *Early Visual Processing in Human Cognition*

As in the toad's visual system, the first elements in the human perceptual system are the cells in the retina. We also have two different kinds of cells in the retina: cones and rods. Each of these cells is at the onset of a different neural path (the dorsal and ventral path) and each one is supposed to provide a different kind of information. Let us concentrate on cones, which start the path that, among other things, computes color.

Cones are photoreceptor cells. There are three different kinds of cones that react to light waves with different amplitude and frequency. Some respond most to light of long wavelengths (L), peaking at a greenish yellow color; others to light of medium wavelengths (M), peaking at green color and the third type responds most to short-wavelength light (S), of a bluish color:

In teleosemantic terms, we could say that every kind of cell has a relational function. For instance, activity in cells of type L maps onto a certain set of waves according to a complex mathematical function, which differs from the mapping function of activity in S and M cells. Since the mapping function of every kind of cone is different, activity in every kind of cell generates a different sort of representation. When two cones of different kind (say, L and M) fire at the same rate, they represent a different feature, because they are supposed to correlate with different states of affairs. This is precisely the reason evolution has endowed us with different kinds of cones. Roughly, the idea is that

12 On the other hand, notice that even if I added all we know about perceptual systems, the account would still be severely incomplete due to the significant gaps in our current scientific knowledge.

every cell should be considered a small representational sub-system that generates certain kind of information.

Similarly, each cell might have different degrees of activation. Since firing always has the same intensity, by 'cell activity' people usually mean the *frequency* in which a a neuron 'fires'. The more frequent a cell fires, the more active it is. So not only different kinds of cells produce different representations, but the same kind of cells firing at different rates generate a representation with a different content. If we also consider the fact that each cell has a different receptive field, we have a rough description of the complex kind of representation that is generated by the activation of the retina at a certain time.

4.2.3.2 *Two-path hypothesis*

In order to explain how further computations take place in human perception, we need to take into consideration the widely recognized fact that what we call 'the visual system' is actually composed of two different systems (Milner and Goodale, **?**; Jacob and Jeannerod, 2003; Farah, 2000; Raftopoulos, 2009aa, 2009bb; Millikan, 2004, p.175-7, Berenthal, 1996). On the one hand, the dorsolateral pathway (which roughly begins at the occipital area and runs along the parietal lobe) generates visuomotor representations, which are closely tied to (lower-order) performances like reaching or grasping (Jacob and Jeannerod, 2003). The area that is in charge of producing these representations is sometimes called the 'where', 'how' or 'pragmatic' path (Milner and Goodale, **?**). This dorsolateral path should be clearly distinguished from the ventral pathway, which is mainly located in the temporal lobe. The ventral pathway is the visual path responsible for our conscious representations and most scientists agree that its primary function is recognitional. It is sometimes called the 'what' pathway (Milner and Goodale, **?**). Furthermore, this path is also involved in higher-order performances, like the representation of distal goals.

The evidence available (which includes similarities between brain structures of other organisms -Milner and Goodale, **?**) clearly suggests that the two paths are functionally, locally and neurologically differentiated, even if there are many interactions between them. The most striking evidence in favor of this claim comes from cases that illustrate a 'double dissociation', that is, cases that show that one of the systems can work more or less normally, while the other is severely impaired. The most famous case, presented in Milner and Goodale (**?**), was patient D.F., who could not identify many objects but was able to interact with them quite satisfactorily (Jacob and Jeannerod, 2003). On the other hand, some other patients with different lesions keep their recognitional abilities, but suffer from serious motor impairments in grasping (Carruthers, 2000). These sorts of experiments have been replicated with other patients, so that nowadays the evidence for the dual-path hypothesis is strong enough for supporting a wide consensus within the scientific community.

In order to show how neuroteleosemantics can be employed in these cases, let us sketch very briefly how representations are produced and computed in every path. This is of crucial importance for developing a teleosemantic account of concepts.

VISUOMOTOR REPRESENTATIONS Let us start by investigating what kinds of representations are encoded in the visuomotor system (that is,

the dorsolateral pathway). Notice that, since these representations are not conscious, we cannot use introspection as a method for determining its content.[13]

The first question is whether the visuomotor system represents the presence of something like objects. Evidence gathered over the years suggests that both streams may at least represent some degree of objecthood (Sholl, 2001). For instance, according to Pylyshyn (1999, 2003, 2007) objects (or, rather, things or proto-objects) are already represented at early stages of the visual system. He brings forward a set of MOT (Multiple Object Tracking) experiments, in which subjects are asked to follow four objects on a screen and these objects continually change their color, size, location and any other property that may be used to identify the thing. Since subjects are capable of following these objects and recognize them as the same thing, Pylyshyn concludes that the mental states deployed in resolving this task (which he calls FINST for 'FINger INSTantiation') work as pointers and represent the presence of certain objects or proto-objects (Pylyshyn, 2004). So something like 'thing' or 'objecthood' may be represented in early stages of the visual process. Other experiments involving priming point in the same direction (Scholl, 2001, p. 7). Consequently, since the dorsolateral and ventral pathways differ only after the processing that takes place at the V1 and V2 regions, and Pylyshyn's FINSTs are located before these two paths diverge, this evidence supports the view that both pathways represent the presence of things or proto-objects (see also Raftopoulos, 2009a, p. 110).

Furthermore, it is also well-known that the dorsolateral pathway computes (among others) information about size, distance and spatial information (Frisby and Stone, 2010; Pylyshyn, 2003, 2007, Raftopoulos, 2009aa, p. 111), although some evidence suggests that the cortical region that processes information about space might be to some extent functionally and anatomically differentiated (Jacob and Jeannerod, 2003). In any case, it seems that the spatial information is represented in this pathway as being subject-centered rather than object-centered. This constitutes a crucial distinction between the information encoded in the dorsolateral pathway and the information encoded in the visuomotor pathway (Farah, 2000).

If we put all these data together, we get the following picture: there is a whole range of computational processes taking place in different parts of the brain and, at some point, a complex representation in the dorsolateral pathway is produced, which represents something like *there is an object [or thing] with such size, shape, at such subject-centered location,...* This complex representation is the outcome of a large number of representational states produced and consumed by different systems in chain. The result of this set of computational steps in the sensory system is a complex representation with this content. Now it is time to apply the third step in Procedure. Is it plausible at all that the existence of this state of affairs is what the consumer-system needs in order to perform its other functions Normally?

Let us start by asking the most simple question: what is the consumer system of these representations? Scientists suggest that the the dorsolateral pathway is mainly involved in reaching and grasping actions as well as (more or less) automatic responses. That claim is supported with

13 This is not to say, of course, that introspection is a reliable method for discovering the content of our mental states (see Pylyshyn, 2007).

behavioral as well as neurophysiological evidence based on the fact that the dorsolateral pathway connects with the frontal lobe, which is partially devoted to perform motor functions. As Raftopoulos suggests:

> The dorsal stream utilizes visual information for guidance of action in one's environment. For that purpose, *it needs* to have information about the dimensions of objects in body-centered terms, that is, in an absolute frame of reference (...) Size, for instance, is computed in an absolute metric, that is with respect to the perceiver, and not relationally with respect to the sizes of other objects in the scene. *To see why that has to be so*, recall that the dorsal system subserves an organism's on-line interaction with the environment. *Successful action requires* that, say, the size of a body be perceived and acted upon in an absolute metric and not in a metric that relates it to other objects in the scene. *To grasp successfully an object, one needs* to perceive its real or absolute size, so that the aperture of the handgrip fits the real size of the object not its relational size. (Raftopoulos, 2009a, p. 110-1. Emphasis added)

In these quotations, I have stressed those parts where Raftopoulos assumes that certain properties are represented precisely because this is what its consumer systems requires. Hence, it seems that representing the presence of objects with a certain absolute size, at certain ego-centrically determined location,... is precisely what the consumer system needs for successful action. These are the kind of properties that are required for the motor system to act appropriately. The color of the object or its identity do not seem to be as relevant for reaching and grasping actions as its size and ego-centrically determined location, so the result seems to fit with our assumptions. Therefore, it seems that, at least in this case, the discriminatory capacities of the elements in the dorsolateral pathway and the needs of the systems that consume these representations yield the same result (i.e. step 3 in PROCEDURE does not apply, see 4.1.4.2). Again, this is a case where the methodology proposed at the beginning seems to provide the right results.

Furthermore, I think that this conclusion lends support to the strategy of assuming that quite often the properties a complex cognitive organism is able to discriminate and the properties that are relevant for action are likely to fit together very nicely. As I said, preference must be given to the needs of the consumer system since that follows from our theoretical assumptions developed in part I, but in many cases I think both perspectives will yield the same result. So we do not need to see them as opposing strategies, but complementary ones.

VISUAL REPRESENTATIONS Since visual representations (the ones located in the ventral pathway) produce (or, maybe, are identical to) our conscious experiences, scientists usually appeal to introspection in order to identify the kind of properties that are represented (even if sometimes that might mislead us). Our experiences together with evidence from cognitive science shows that the ventral pathway involves representations of colors, relative size (what, among other effects, is responsible for the Müller-Lyer illusion), location in an object-centered framework, and many others (Farah, 2000). Here 'the information is represented in a relational frame of reference in which objects are

represented in a scene and not with respect to the body of the perceiver' (Raftopoulos, 2009a, p.110).

Since both phenomenological evidence and cognitive science tells us that there are representations of all these properties in the ventral pathways, I think we can tentatively conclude that such representations exist and that the content of visual states is something like *there is an object [or thing] at such object-centered location, with such color, relative size,...* Crucially, in contrast to the dorsal pathway, features are represented here within a relative or relational frame of reference (with respect to other objects). Size, for instance, is computed in relation to the size of other objects. Evidence for the the contrast between object-relative size represented in the ventral pathway and the absolute size encoded in the dorsal pathway comes primarily from several illusions (like the Ebbinghaus illusion) involving several objects that fool the ventral system but seem to leave the representations in the dorsal stream unaffected (Jacob and Jeannerod, 2003).

In the case of the ventral pathway, however, there is an additional difficulty. We saw that visuomotor representations are mainly consumed by a system responsible for reaching and grasping actions, and in this case we saw that the content we obtained using Relative Indication* corresponded with the needs of the consumer system. However, visual representations feed (at least) two consumer systems. That means that, when assessing the representational content of states in this visual pathway, we need to consider the needs of two different systems.

On the one hand, representing relative position, object-centered location, etc... is very useful for reaching and grasping actions, in the same way as visuomotor representations. Certainly, subjects that have suffered lesions in the temporal area might still be able to reach and grasp objects, but it is pretty clear that this ability is also impaired in important respects. Consequently, I take it that the representations in the ventral pathways also contribute to these fairly automatic responses.

Nonetheless, the main function of perceptual representations is not to guide immediate action but to enable recognition of the objects. It is by means of having a (conscious) visual representation of a certain object that we manage to recognize it as the same object we have been gathering information about. The capacity of recognition enables us to use previously gathered information to this new situation and hence to improve the chances of acting appropriately. Therefore, allowing us to reidentify objects and use this information in order to shape our actions is the main function of visual representations (in the sense of 3a of Third Teleosemantics, see3.2.5)

Now, following our methodological strategy, it seems that the kind of content that results from taking the perspective of discriminative capacities (something like *there is an object [or thing] at such object-centered location, with such color relative size,...*) is in accordance with the needs of the consumer systems, which feed into the motor and recognitional system. Hence, again, this is a case in which our methodology and assumptions have proved to be helpful and illuminating.

Finally, it needs to be said that I have been focusing on visual perception, since this is the area that has centered most attention in philosophy and cognitive psychology, but I have been assuming that the kind of explanation offered here could be extended to other perceptual systems, such as hearing or smelling. If a computational approach is also true of them, it seems that in principle the same analysis in terms of a set of

sender-receiver structures instantiated in different relays could also be provided. Thus, I will assume that the account I have provided yields a teleosemantic account not only of visual perception, but an account of perceptual states.

4.2.3.3 *Perceptual Tracking*

So far in this chapter I have explained how the teleosemantic approach developed in part 1 could be satisfactorily applied to perceptual states in the visual system, which obviously is an interesting issue on its own. Nevertheless, the result we obtained from the previous discussion might be relevant in order to define a process that will have an outstanding relevance in the chapters to come: perceptual tracking.

Perceptual tracking is (roughly) the capacity of representing a given entity as the same through temporal and spatial changes. The ability of tracking an entity involves the perceptual states we have been discussing so far and probably many other mechanisms.

As a first approximation, it seems that a subject tracks an entity only if this entity satisfies the descriptive content of the perceptual state. For instance, if John is perceptually tracking an object o while having an experience with the content *there is a shiny and large red object at location l*, then the object o must instantiate (at least many of) the properties represented. A subject can be said to be perceptually tracking an object o only if o happens to satisfy (a sufficient number of) these attributed properties.

So, we could tentatively define the ability of perceptually tracking an object in the following way:

FIRST TRACKING A subject A perceptually tracks a particular entity e at $t_{1...}t_n$ only if e satisfies (to a certain degree) A's perceptual content at $t_{1...}t_n$

Of course, satisfying some properties might be more important than satisfying others. Location, for instance, seems to be more important than color. I can be perceptually tracking an entity E even if I am radically misrepresenting its color, but it seems that I cannot be perceptually tracking an entity if I am radically mistaken about its location (Evans, 1982). In any case, whether a subject tracks, perceives or misperceives an entity seems to be a matter of degree.

Now, I spelled out this principle in terms of necessary (and non-sufficient) conditions because there are two central features missing in this analysis of perceptual tracking The first one concerns attention. Attentional mechanisms have increasingly been studied in recent times, and there is still a vivid controversy concerning its analysis (Pylyshyn, 1999, 2003, 2007; Raftopoulos, 2009a). In that respect, one might worry that a necessary condition for tracking an entity involves the subject attending to it. This is probably right, so FIRST TRACKING will have to include this mechanism.

Nonetheless, while I think attentional processes should be added into the definition, I doubt there is any prima facie reason for thinking that they can pose any problem to the teleosemantic perspective offered here. Attentional mechanisms (which are usually divided into 'location-centered attention' and 'objected-centered attention', Raftopoulos, 2009a) should be interpreted as further structures that yield the perceptual content that figures in the right-hand side of the definition. It is usually thought that they modify perceptual content either by

filtering out some represented features or by reducing the threshold that is required for a certain perceptual mechanism to fire (Raftopoulos, 2009a). In principle, it seems both of these processes can be accommodated within the framework offered here. Consequently, they should be analyzed as part of the complex mechanism that contributes to the production of a perceptual representation. We can simply add this element within the definition of tracking:

SECOND TRACKING A subject A tracks a particular entity E at $t_{1...}t_n$ only if

1. E satisfies (to a certain degree) A's perceptual content at $t_{1...}t_n$
2. E is being attended by the subject.

On this modified version, a subject tracks a certain entity at a given time only if this entity satisfies the (complex) perceptual content of the subject's experience and that entity is being attended to by the subject (for a similar claim, see Dickie, 2010, p. 228). Of course, this is a very superficial characterization of the perceptual and attentional mechanisms, but as I said, my main purpose here is not to provide a full description of perceptual mechanisms, but only to suggest that the sender-receiver teleosemantic model I put forward in the first part of the dissertation could be applied to perception.

The final feature that needs to be added is that the subject must somehow recognize that the object at $t_{1...}t_n$ is *the same object*. After all, even if I am attending and perceiving a certain object o during a period of time, we would not claim that I am tracking the object unless I somehow presume that the object I am perceiving at different times is the same. Following Recanati (Forthcoming) and Millikan (2000), I think it is better not to conceive this act of identification as the subject entertaining a further identity thought of the the form A=B, what seems to be empirically inadequate. Arguably, the best way to cash out this idea is in terms of the subject being disposed to treat both signs as referring to same entity. In a nutshell, the idea is that in order to track an object o a subject must be disposed to react as if the object she is perceiving at $t_{1...}t_n$ were the same. If we add this idea to the definition, we get a more plausible set of necessary and sufficient conditions for tracking an object:

BETTER TRACKING A subject A tracks a particular entity E at $t_{1...}t_n$ iff

1. E satisfies (to a certain degree) A's perceptual content at $t_{1...}t_n$
2. E is being attended by the subject.
3. A is disposed to behave as if the entity it is perceiving at $t_{1...}t_n$ was the same.[14]

As we will see in 6.3.2, the ability of perceptual tracking underlies the capacity to recognize objects in perception and to gather and use the information acquired in this process. Furthermore, perceptual tracking is the key process that underpins our capacity for conceptual tracking,

14 This account satisfies what Pylyshyn (2004, p. 804) calls the Discrete Reference Principle, according to which in order to track an object (1) each individual object in that set must be kept distinct from every other object in the display and (2) each individual target object must be identified with a particular individual target object in the immediately preceding instant time.

and this is precisely the sort of ability that grounds our capacity for conceptual formation. These ideas will be developed in more detail at the end of chapter 5 and specially in chapter 6.

4.2.4 *Decoupled Representations*

Finally, I would like to show why the account I have presented so far overcomes the third problem set up at the beginning of this chapter: the question of decoupled representations (see 4.1.2).

The problem was the following. One might argue that, according to the teleosemantic framework offered in part I, the content of a representation (what a representation is supposed to map onto) is determined by the needs of the consumer system. So in order to determine the content of a representation, we must know the needs of the system that consumes these representations. Now, perception is a system that produces representational states, but it seems that there is no specific job these states are supposed to help to bring about. Our perceptual states do not force us to produce any particular effect; on the contrary, they enable us to perform a wide range of activities, and hence it is not clear how the sender-receiver structure applies here.

I think there are two possible replies to this sort of worry. On the one hand, it can be argued that there is an effect that always ensues perceptual representations: they prepare the organism (in a very specific way) to do certain things. For instance, a particular species of Australian jumping spider (*Portia fimbriata*) which preys on other spiders has a set of chemoreceptors sensitive to the silk of another family of spiders, *Jacksonoides Queenslandicus*. When a jumping spider senses the silk it "lowers the thresholds for responses by central nervous system modules (or feature detectors) associated with the visual system" (Harland, 2004, p. 37). In other words, the presence of silk makes the jumping spider more sensible to visual cues, and hence enables it to detect a prey more easily. Crucially, there is no particular *behavior* that the chemoreceptor system is supposed to elicit; nevertheless, there is always a particular (*internal*) effect: it changes the internal configuration of the organism, such that it is ready to react in the appropriate way to different circumstances. One might argue that, in a similar fashion, the effect of perceptual states may consist in putting the organism in a state that enables it to react appropriately towards different scenarios. Perceptual representations always have the effect of putting the organism in a certain dispositional state.

There is a second possible reply to the worry of decoupled representations, which is compatible with the previous one. The point is that, when properly understood, the existence of decoupled representations does not clash with the key tenets of teleosemantics. Teleosemantics does not require that the representation elicits a particular effect. Let me illustrate this point with an example. Male and female grasshoppers produce an acoustic signal in order to inform members of the opposite sex about their respective position (Staaden, 2004). As a matter of fact, when a male hears the signal of a female they usually approach each other in order to mate. Now, let us imagine a similar species, call it grasshopper*, in which males, upon hearing a sign emitted by a female, have three actions available. First of all, if male members of this imagined creature are ready to mate, male grasshoppers* still approach the female grasshopper*. If they are willing to mate, but

are exhausted (if they had a really busy day), they just respond back and await the approach of the female grasshopper*. Thirdly, if male grasshoppers* are not willing to mate, they inform other males that there is a female willing to mate. Notice that all these reactions require the same state of affairs in order to be successful: the presence of a female grasshopper* willing to mate. Additionally, we can imagine that a grasshopper usually has to reflect on its internal feeling, in order to know whether it is tired or willing to mate. So, the response is not automatic, but mediated by reflection. The key point of the example is that there is no direct and automatic link between the signal emitted by a female and the behavior of the male grasshopper*, so it is a decoupled representation. And, nevertheless, since all these varied behaviors require the same circumstance in order to be successful (the presence of a female grasshopper* willing to mate), this is the state of affairs represented by the female signal, according to teleosemantics. As a result, the existence of decoupled representations is entirely compatible with the story I have given so far.

Of course, perceptual systems are even more complex than the signaling system of grasshoppers*, because the former have open relational functions (see 3.2.2) and, consequently, they can produce new representations (e.g. perceptual representation of red_{476}). But, concerning the problem of decoupled representations, the same explanation holds: in perceptual systems, the fact that the percept maps onto the real world according to certain mapping functions explains why a wide range of behaviors were successful, in the same way that the fact that the grasshopper's* mental state maps onto the presence of a female grasshopper* explains the success of a set of different behaviors. Put it in a different way: your perceptual state does not force you to perform any particular action, but the fact that it always represents the external world in the same way (according to the same mapping rules) explains the success of any action you perform with its aid. The latter is the only effect that is required by teleosemantics.

Indeed, there is a feature of perceptual systems that suggests that this account is on the right track. When a representation is coupled and hence always elicits an automatic behavior, the representation does not need to have parts that correspond to parts of the representatum (i.e. it can be inarticulated) and its physical format can be extremely simple. In contrast, if a representation has to be used for many purposes, there is a strong tendency to represent more features (which might be useful for some behaviors and not useful for others) and to be more complex. Thus, we can expect decoupled representations to be much more complex (semantically and syntactically) than coupled representations. An analogy might be useful here: if I tell my mechanic that my car is broken because the carburetor is not working, he will take a very particular set of tools to fix it; if, instead, I tell him that he must be prepared for any kind of reparation, he will have to take a wider set of tools. In a similar fashion, the fact that there is no specific behavior decoupled representations are supposed to bring about suggests that they must represent a wide variety of features.

That prediction fits perfectly well with the description of perceptual systems which I have offered in this chapter: perceptual states are extremely complex and represent a huge variety of features precisely because they can be used to perform an extremely wide range of tasks. But the semantic and syntactic complexity of perceptual systems is a

striking feature that I just argued that can be partially explained by the theory suggested here.

In conclusion, I think that the fact that perceptual representations are decoupled in the sense of Sterelny (2003) should not be regarded as a problem for the application of the teleosemantic framework.

4.2.5 *Conclusion*

It is time to take stock. We saw that none of the preliminary difficulties outlined at the beginning threatens a neuroteleosemantic approach to perception and cognition. Indeed, I set up a general methodological strategy that is supposed to fit with the way neuroscientists proceed and might help us to investigate how the sender-receiver structures and representations are actually instantiated in the brain.

After showing how Third Teleosemantics works for computer models and toad brain states, I argued that it could also be employed in the two visual pathways that compose our visual representations. In that respect, I argued that, in both pathways, the cues that satisfy Relative Indication* are precisely the properties that the consumer system seems to need. So, in the examples discussed, the methodology described at the beginning of the chapter (Procedure) was vindicated. The stimulus that most strongly elicits the representation and the needs of the consumer system often seem to fit each other nicely.

In the final part of this chapter, I would like to discuss how the teleosemantic model I have developed relates to current discussions in the philosophy of perception.

4.3 philosophy of perception

In the first part of this chapter I have surveyed the psychological evidence on toad and human cognition and I have shown how the teleosemantic proposal can be applied to this domain. The resultant picture has strong philosophical implications. In this section, I would like to examine and discuss some philosophical arguments that might threaten the view of perception I presented in this chapter.

There are three consequences of the proposal outlined here I would like to address: (1) the discussion whether perceptual states are representational, (2) the debate between conceptualism and non-conceptualism and (3) the question whether perceptual contents are singular or general.

4.3.1 *Do experiences have content?*

First of all, one might worry that the perspective I offered on the contents of perception trivializes certain debates. For instance, an enduring discussion in the philosophy of perception concerns the question whether perceptual states are representational. My account has assumed (and lend support to) the view that experiences are endowed with representational content. As I have already pointed out in several places of this dissertation, the neuroteleosemantic account suggested here assumes that the same sender-receiver model can be instantiated in simple and basic representational systems as well as in sophisticated cognitive abilities. And since I have arguing that bacteria, frogs and

beavers produce representations, it should be obvious by now that the perspective defended here entails that perceptual states are also representational.

Even if this sort of minimal representationalism is widely held to be true (Peacocke, 1992; Chalmers, 2006; Pautz, 2008; Tye, 2000,2009a; Schellenberg, 2010, 2011; Siegel, 2011; Burge, 2010) it is not completely uncontroversial (Campbell, 2002; Martin, 2002; Fish, 2010). So before moving ahead, let me just point out some of the compelling reasons that suggest that perceptual states are endowed with representational content.

The first obvious reason why we may hold that perceptual experiences represent the world as being in a certain way is that it agrees with how things appear to us. I think it is plain that when we are undergoing an experience, it seems to us that our experience is representing the world as being in a certain way. The claim that experiences are *of* certain environmental features has a clear intuitive appeal. The same point can be put in a different way; it is hard to imagine a subject which has an experience that is qualitatively identical to ours, but which does not represent anything. While it might be conceivable, the mere fact that imagining this case strikes us a extremely odd makes a prima facie case for the view that experiences are representational.[15]

Secondly, the commonalities between cases of accurate perception, illusion and hallucination can be nicely explained by this minimal form of representationalism (Tye, 1995, 2000). Notice that cases of veridical perception, illusion and hallucination are different in crucial respects: in hallucination there is no object perceived, in illusion there is an object that lacks some of the properties attributed to it and in veridical perception there is an object that has all the properties one is attributing. And, nevertheless, the three cases might be indistinguishable from the first person point of view. Since the three experiences might share phenomenology, there must be a feature that they have in common and that accounts for these commonalities. Representational content might be such a feature.

Two further arguments are often used in support of minimal representationalism, which have to do with the explanation of some cognitive phenomena. First of all, minimal representationalism can explain why there is some cognitive penetrability between perceptual experiences and higher cognitive abilities (Schellenberg, 2011). For instance, it is widely known that certain beliefs or knowledge can influence the way we perceive the world. E.g. someone who knows German has a different experience when listening to someone that speaks German from someone who does not know this language. Minimal representationalism gives us an easy way of accounting for this fact: since beliefs represent the world as being in a certain way, if we accept that experiences are representational as well, we can explain the phenomen-

15 This point is closely related to the argument for the transparency of experience, which claims that when we try to focus on the phenomenal properties of our experiences, the only thing we seem to be aware of are properties of the object of experience (Martin, 2002; Tye, 2000). Nevertheless, notice that I am not appealing to the transparency claim in order to support this sort of minimal representationalism. The transparency argument heavily relies on the fact that we are not aware of any non-intentional property of our experience; in contrast, the claim that our experiences are representational does not need to make this assumption. Even if we were aware of non-intentional properties of our experience, we could also be aware of their representational properties, so the denial of transparency is still compatible with the intuition I rely on. Alternatively, if one is convinced by the transparency argument, that in itself is an outstanding argument in favor of the sort of representationalism I am advocating.

ological effect by appealing to an influence in content. Finally, minimal representationalism helps us to explain how can we remember experiences. The idea is that when we remember an experience we recover a state with the same content and that is what explains the possibility of such memories and the phenomenological commonalities between the original and the current experience (Schellenberg, forthcoming).

I think all these arguments strongly suggest that we have very good reasons for thinking that perceptual states are representational. Of course, some people deny the claim that perceptual states are endowed with representational content. These views are usually classified under the label of 'Naïve Realism'. Unfortunately, discussing the arguments of Naïve Realists would lead us too far away from our present concerns. My aim was merely to provide some independent support for the claim that perceptual states are representational. Of course, if this thesis succeeds in providing a teleosemantic account of perception and concepts, that would provide an additional argument in favor of the claim that perceptual states are representational.

4.3.2 *Conceptualism*

The second debate that a neuroteleosemantic proposal trivializes is the question on perceptual non-conceptualism. An enduring discussion in the philosophy of perception concerns the question whether the contents of perceptual states are conceptual or non-conceptual. It is common nowadays to distinguish two versions of non-conceptualism, the state-view and the content-view (Heck, 2000, p. 485). On the one hand, state nonconceptualism claims that a subject can (perceptually) represent a certain feature F even if he lacks the concepts for specifying F. On the other, content non-conceptualism claims that the content of (perceptual) representational states is different in kind from that of cognitive states like belief. State and content conceptualism are the denial of these theses.

Now, if we assume the perspective developed in this thesis, it seems that state non-conceptualism is trivially true and content non-conceptualism trivially false. The reason state non-conceptualism is trivially true is that we have been assuming that many organisms that very probably lack concepts (bees, toads, salamanders,..) have perceptual states that represent the world as instantiating many properties. So there are many properties F such that an organism can represent F without having the concept for F. Similarly, I have argued that very many sub-personal states (like neuronal activation in early visual processing) are representations in the full-blown sense. According to my proposal, these states represent many complex features and it is extremely plausible that people do not need to have the concept F in order to represent F in early vision. So it seems that, if we assume the picture put forward in the dissertation, state non-conceptualism is trivially true.

Secondly, content non-conceptualism is trivially false because in this dissertation I am assuming (and, at the same time, arguing for) a continuum between the content of simple representational states and conceptual content. The content of perceptual states is not different in kind from the content of concepts (additional argument were provided in 4.1.3).

I admit that these are consequences of the view defended here, but I do not consider them an unwelcome result. If a well-argued position

makes a certain debate uninteresting, so much the worse for the debate. Indeed, it is noteworthy that there are many views that make the debate between conceptualism and non-conceptualism trivial (Toribio, 2008). Therefore, whereas I acknowledge this is an important consequence of the view depicted here, I do not think it should be regarded as problematic.

4.3.3 *Are the contents of experience singular or general?*

Finally, once it is accepted that experiences have representational content and that this content is state non-conceptual and content conceptual, there are still two main ways of specifying this content. On the one hand, the standard view claims that perceptual states are endowed with singular content, that is, that the content of perceptual states involves the object of perception (Bach, 2007; Burge, 2010; Tye, 2009b; Schellenberg, 2010, 2011; Siegel, 2011). According to this approach, which I will call the 'Singularity View', the content of perceptual states can be appropriately described using an indexical expression like: *That object has such and such properties*. On this view, if at t_1 John is perceiving object$_1$, then the accuracy conditions of his perceptual state involve object$_1$. Similarly, if at t_2 object$_1$is changed for a qualitatively identical but numerically distinct object$_2$, then the content of his perceptual state is different; at t_1 John's perceptual state had the content *that object$_1$ has such and such properties* and at t_2 it has the content *that object$_2$ has such and such properties.* This is said to be so even if the two experiences are indistinguishable from the first-person point of view.

An alternative view, which Tye (2009a) calls the 'Existential View' holds that the content of perceptual states is general, that is, that veridical experiences of qualitatively identical but numerically distinct objects can share content (Davies, 1992). Hence, on the Existential View we can appropriately describe the contents of perception using an existential quantification as follows: *there is an object with such and such properties*.

Now, my own view is that the Existential View is on the right track, and hence I describe the contents of perceptual states as involving an existential quantification. Let me provide some reasons for thinking that this is the right view on the contents of experience.

DEFAULT VIEW First of all, so far we have been assuming that the content of simple representational states seems to be rightly spelled out as an existentially quantified content (*there is nectar at such and such location, there is a fly around, there is a small insect at l,...*). Given that I have shown that the Sender-Receiver framework is instantiated in exactly the same way in simple representational states and perceptual states, the default view should be the Existential View. We need good reasons for thinking that the content of the latter should be specified in a different way.

MISREPRESENTATION Another reason in favor of the Existential View is that it seems to provide a better explanation of perceptual misrepresentations. If the content of perceptual states involved particular objects (that is, if the content of perception was *singular*, as the critics of the Existential View suggest) then in cases of hallucination the content of the perceptual state would be gappy. In other words,

on the Singularity View the content of my perceptual state would be something like __ *has such and such properties*. The main problem with that consequence is that it is not clear that a gappy content of that sort can be inaccurate. And if it is not inaccurate, then the Singularity View entails that hallucinations are not inaccurate representational states. In contrast, the Existential View can straightforwardly accommodate inaccurate representations: since in cases of hallucination there is no object being perceived, then the existentially quantified content of the perceptual state (*there is an object with such and such properties*) is simply not satisfied, and it is therefore inaccurate.

Despite these advantages, critics of the Existential View usually adduce an arguments in favor of the Singularity View: The Particularity Intuition. Let me briefly consider this argument.

PARTICULARITY This argument, which has recently been pointed out by many philosophers, concerns the *particularity* of perception (e.g. see Brewer, 2006; Campbell, 2002; Martin, 2002, Schellenberg, 2010). Basically, the idea is that perceptual states seem to involve a relation to particular objects. My perceptual state of an object A seems to be of *that* particular object. According to some people, since it looks to me as if my perceptual experiences were directed at particular objects, the accuracy conditions of perceptual states should reflect this fact. Consequently, they argue, the content of my perceptual state should be singular.

While it is undeniable that those philosophers that appeal to the particularity of perception are trying to make an intuitive claim about our experiences, a serious difficulty with the particularity objection is that it is extremely difficult to come up with a precise formulation of the intuition that is both plausible and at the same time supports the Singularity View. Several definitions that can be found in the literature turn out to be unsatisfactory. In particular, there are four common ways of spelling this idea out that fail to lend support to the Singularity View over the Existential View:

(Satisfaction) Soteriou (2000) spelled out the particularity intuition in the following terms:

> If an experience is a perception, then the experience has particularity. There is some fact that determines which particular object is *represented* by the subject's experience. (Soteriou, 2000, p.178. Italics from the original)

If all we mean by particularity is that in every successful occasion of perception there is some fact that determines which particular object makes my perception true, the Existential View does not fail to fulfill this desideratum, since in every occasion in which $\exists x Px$ is true there will be a particular object that satisfies this description. Supporters of the Existential View may accept that there is always some fact that determines which particular object is represented by the experience. So, if the particularity objection is cashed out in this way, it fails to lend support to the Singularity View over the Existential View.

(Individuation) In a recent paper, Schellenberg (2010) points at the existence of a relational particularity, according to which the object of

experience should play a role in the *individuation* of the experience and she claims that the Existential View cannot accommodate this fact.

However, it is close to a platitude that experiences can be individuated in many different ways; thus, if the particularity claim merely states that the object of experience should play a role in individuating the experiential state, it is not clear that the Existential View fails to satisfy this desideratum. For instance, one could endorse the Existential View and account for that role of objects in the individuation of mental states by appealing to the relations of causation or satisfaction. That is, the Existential View can perfectly assume that the object I am causally related to or the object that satisfies the experience's existentially quantified content plays an important role in individuating the experience. If, on the other hand, experiences are individuated by phenomenal character, the Singularity View and Existential View are on a par. Consequently, if particularity is spelled out in terms of individuation, then this objection has no bite.

(Content) Some people argue that the particularity intuition consists in the claim that the *content* of perceptual experiences should be object-involving (Schellenberg, 2010). On that interpretation, the particularity intuition boils down to the following: the thesis that the content of experience is object-involving is intuitively compelling. Obviously, the problem with this way of formulating the argument is that the particularity intuition was supposed to lend support to the Singularity View, but under this interpretation the particularity claim turns out to be a mere notational variant of content disjunctivism. Content disjunctivism is precisely the view that the content of experience is object-involving, so if the particularity intuition is cashed out in this way, it would not provide any additional support for this approach. The particularity intuition would amount to the thesis that content disjunctivism is intuitive, what of course does not bring anything new into the debate.

(Looking) A last proposal is to interpret the particularity intuition as a claim about the way things look to the subject. The idea, then, would be that the accuracy conditions of the perceptual state should include the object because it *looks* to the subject as if she was perceiving a particular object.

However, I doubt any formulation in terms of looking can lend support to the the Singularity View. First of all, it is not obvious that the way things look can decide between different ways of specifying the content (what is the kind of look that a state with an existentially quantified content is supposed to have?). Secondly, if we accept the plausible view that perceptual states and hallucinatory states can look the same way and given that the Singularity View is committed to the claim that these two kinds of states have different kinds of content (one is object-involving and the other is gappy), this theory is committed to accept that in some cases at least how things look does not determine content. So it is not obvious that the Singularity View can hold a strong connection between the way things look and the state's content. In contrast, the Existential View is compatible with a strong connection between ways of looking and content, because it can coherently hold that perceptual states and hallucinations share ways of looking and content.

In conclusion I think that the particularity objection fails to provide any independent support for the Singularity View and against the Existential View.

Despite the fact that I think that the Existential View on the contents of perception is more plausible than the Singularity View, it is important to point out that, in principle, teleosemantics is not incompatible with the latter. So, even if I have argued the Existential View fits better into the teleosemantic framework, a teleosemanticists could coherently hold that the content of perceptual states is singular. If one holds the Singularity View, she would only need to slightly modify some definitions (such as Perceptual Tracking) accordingly. No major issue would be affected. Consequently, in the rest of the thesis I will keep talking as if the content of perceptual states could be satisfactorily specified with an existential quantification, but if one thinks that perceptual contents are singular, that is indeed compatible with anything I will say in the remainder.

4.4 conclusion

In this chapter, I have shown how a neuroteleosemantic account of cognition can be developed. I have carefully described how the tools we devised in the first part of the dissertation can be applied to perceptual systems in toads and humans, even if, of course, a full characterization of every step in the formation of the percept lies beyond the scope of this thesis. In the final section, I have stated some consequences my view has on certain current debates in the philosophy of perception.

The task of the last two chapters is to employ the teleosemantic framework outlined in part I and the naturalistic account of perceptual content sketched in this chapter in order to naturalize a central cognitive ability: concepts.

5 CONCEPTS

If in the last chapter I showed how the teleosemantic framework outlined in the first part of the dissertation could be applied to the perceptual system, in these two final chapters I would like to focus on a different but equally fundamental cognitive ability: concepts. The task of the remainder of the dissertation is to provide a naturalistic (teleosemantic) theory of the semantic properties of concepts.

This first chapter on concepts has two main goals. First, I will describe in some detail the entity we are trying to naturalize, that is, concepts. This preliminary discussion is crucial because concepts have been defined in many different and incompatible ways. Furthermore, theories disagree on the kind of semantic properties concepts are endowed with, so the naturalization of semantic properties would constitute a different project depending on the way we approach concepts. There are three aspects in which the notion of concept has to be clarified: its nature, its structure and how its content is determined. I will seek to disentangle these different questions.

Secondly, I will examine some recent attempts to naturalize the content of conceptual representations. I will argue that most naturalistic accounts fail for one of the three following reasons. First, some people (mainly psychologists) assume that one can naturalize the content of concepts by merely appealing to some kind of law-like connection between concepts and their referents (sometimes by explicitly appealing to some of the views discussed in chapter 1). Secondly, most current philosophical accounts fail because people with teleosemantic inclinations think that once we have achieved a set of tools required for a naturalization of content for simple representational states, applying them to concepts is straightforward. I will show that this is far from true. Concepts (and other higher-order cognitive abilities) pose new and difficult challenges for any naturalistic theory. Thirdy, I will argue that one of the main difficulties of current approaches is that they do not correctly apply teleosemantics to the perceptual system before developing a metasemantic theory of conceptual content. This is the reason the theory put forward in the previous chapter is so important. My positive naturalistic account of concepts will be provided in chapter 6.

5.1 DEFINING CONCEPTS

The first thing required in any serious investigation on any entity is to get clear about the object of research. In that respect, there are at least three different questions about concepts that must be addressed: the question about its *nature*, the question about its *structure* and the question about its *content.* Even if the goal of this dissertation is to provide a naturalistic account of conceptual *content,* issues about the nature and structure of concepts have important implications for any theory of content. Some of the relations among these three aspects (and the confusions they have originated) will be highlighted and discussed in this first part of the chapter.

5.1.1 *The nature of concepts*

What is the nature of concepts? What kind of entity are concepts? Three main answers have been offered in the literature (Margolis and Laurence, 2011). The first one claims that concepts are mental representations (or brain structures), the second asserts that concepts are abstract objects, and the third view takes concepts to be abilities. In this section I will shortly describe these different ways of thinking about concepts. I hasten to add that I I will not attempt to argue that any of these notions is superior to the others; all three approaches are probably valid and might be useful for different philosophical enterprises. Nonetheless, I will present some reasons suggesting that the view that concepts are mental representations fits much better into our project. So this is the notion of concept I will be using in the rest of the dissertation.

Let me then very briefly outline how concepts are understood in each tradition, and argue why a certain understanding is preferable for our naturalistic project.

5.1.1.1 *Concepts as mental representations*

Some people take concepts to be a certain kind of brain states endowed with representational content. On this approach, concepts are the *vehicles* of thought and other propositional attitudes. More generally, they are the brain states employed when a subject entertains a thought, a hope, a desire and so on. This the standard view in psychology (Murphy, 2002, p. 5; Carey, 2009; Machery, 2009; Prinz, 2002) and it also very popular in philosophy (Fodor, 1998, 2008; Lawrence and Margolis, 2011, Carruthers, 2006). While psychologists and philosophers often disagree on many central properties of concepts (see 5.1.2 below), they typically assume that concepts are mental representations. This is the common ground that underlies many debates in philosophy and psychology.

One reason for the popularity of this approach is that it fits very nicely into the Representational Theory of Mind (RToM). Roughly, the RToM claims that thinking occurs in an internal system of representation, according to a certain set of transformation rules. Beliefs and other propositional attitudes enter into mental processes as symbols and have certain structure (Sterelny, 1990; Ryder, 2009). In particular, beliefs are composed of more basic representations: concepts. Thus, on this view concepts are the elements that compose propositional attitudes. People working within the RToM usually assume that thoughts have a sort of language-like structure, were concepts play the role of the lexicon (although, of course, it is usually admitted that the parallelism between language and thought is not perfect).

One of the virtues of the RToM is that it can explain very well the productivity of thoughts, i.e. the fact that human beings can entertain an unbounded number of thoughts (see 3.2.4 and 6.5.1). If concepts are mental representations, it is easy to explain how they can compose in many different ways, giving raise to a (potentially) infinite set of well-formed beliefs. Furthermore, the RToM plays a crucial role in accounting for how mental processes can be both rational and implemented in the brain (Rey, 1985, p. 237; Lawrence and Margolis, 2011). These and other reasons explain why the RToM has been very popular during the last decades.

Nonetheless, despite of this tight connection, it is important to stress that the view that concepts are mental representations is compatible with the denial of the RToM. Connectionist models of the mind can also use the notion of 'concept' in order to refer to the vehicles of representations (Machery, 2009, p.13). In general, assuming that concepts are mental representations should not commit one to any particular view on the structure of these representations, not even to a language-like structure (Eliasmith, 2003, p. 132). The key tenet is that concepts are conceived as brain structures that represent the world. Whether these brain structures are thought to be mental symbols or complex networks is not essential and will not be discussed in this thesis.

In contrast, the notion of concept as mental representation is obviously incompatible with the view that there are no mental representations (O'Regan and Noë, 2001). In this dissertation I will assume (in accordance with most cognitive science) that brain states represent the world as being a certain way, so I will assume that non-representationalism is false. Otherwise, the project of naturalizing representational content would not make any sense (at least, in the way it is conceived here). Some additional arguments for the view that non-representationalism is unconvincing were provided in 4.1.3.

Finally, let me mention a crucial issue that will be discussed in more detail in the next chapter. There is a common ambiguity in the use of the notion of 'concept' within this tradition. One the one hand, 'concept' is supposed to refer to a mental *state,* that is, an event (or part of an event). When I have an occurrent thought such as TREES ARE GREEN, my brain is activated in a certain way, and a concept is characterized as a proper part of this brain state (say, the part that corresponds to 'TREE'). On the other hand, concepts are sometimes conceived as brain *structures*, that is, as an enduring mechanism contained in the brain. Activation of the conceptual *structure* gives raise to conceptual *states*. This ambiguity is usually harmless, but conceiving concepts as states or as structures that elicit certain states has striking consequences for a teleosemantic account of them (e.g. we saw that attributing functions to states or systems yields extremely different results). For the time being, I will not try to resolve this issue; for the rest of this chapter I will keep using 'concept' in order to refer to states. I will take back this question in chapter 6 (see 6.2.2.1).

5.1.1.2 *Concepts as Fregean senses*

Alternatively, some people identify concepts with constitutive parts of propositional contents. That is, concepts are not regarded as the vehicles of thought (mental states), but as parts of the representational content. Since propositional contents are usually thought to be abstract objects, on this view concepts are conceived as abstract objects, as opposed to mental objects (Peacocke 1992, Zalta, 2001). Concepts are the constituents of abstract representational contents, which are standardly conceived as structured entities (Burge, 2007, 2010).

In this tradition, concepts are usually identified with Fregean senses, because senses and concepts are more discriminating than referents and because both seem to play the same role in cognition. Consider the concepts CICERO and TULLY. It seems plausible to hold that there is a sense in which the thought that CICERO WAS ROMAN and the thought that TULLY WAS ROMAN are different thoughts. Nevertheless, they refer to the same entities, so their difference in cognitive value

cannot merely depend on their referential content. Many people argue that CICERO WAS ROMAN and TULLY WAS ROMAN differ in *the way* they refer to the same person; following Frege (1892) they call these ways of referring to certain entities *senses* or *modes of presentation*[1]. So, according to these philosophers, the *way* of referring to an entity is an aspect of representational content (Peacocke, 1992). Secondly, these ways of referring (senses) are structured entities, that is, they have parts. 'Concept' refers to the parts in which senses are structured. Consequently, on this view concepts are abstract entities, more fine-grained than referential content, and constitute an important aspect of the representational content of mental states.

5.1.1.3 *Concepts as abilities*

A third group of philosophers argue that concepts should primarily be considered as certain kind of abilities (Kenny, 2010; Dummett, 1993; Millikan, 2000[2]). According to them, a cognitive agent has a certain concept when she is able to do certain things. For instance, when she is able to use words in a certain way, discriminate certain things, or reidentify certain entities. The key assumption common to all these approaches is that a subject masters a concept when she has an ability to do certain things. Concepts are primarily conceived as an ability of a special sort. This view traces back to the late Wittgenstein and his anticognitivist arguments (see Kenny, 2010, p.112; Dummett, 1993, p. 98) and used to be very popular, even if nowadays few philosophers or psychologists endorse it.

It is worth stressing that, although in this tradition concepts are primarily identified with certain abilities, supporters of this view usually accept that these sort of abilities actually require those mental states that people of the first group identify as concepts (Kenny, 2010, p. 106-7, Millikan, 2000, p. 2). In other words, while they think concepts should be primarily identified with certain abilities, they accept that that they are closely related to certain mental states that some other people label 'concepts'. Therefore, most of what people in this tradition say is compatible with the claims of those who take concepts to be mental representations or senses. The only disagreement (which is of course important for our discussion) concerns what does 'concept' primarily refer to.

5.1.1.4 *Concepts in a Naturalistic Project*

As I said, in what follows I will interpret 'concept' as referring to mental representations (or brain structures). There are three main reasons for this choice. First of all, in the first part of the dissertation I have offered a naturalistic account of representational *states*. Nothing has been said about abilities or Fregean senses. From the onset, the project was to naturalize the content of certain states and brain structures, so I in what follows I will also be concerned with concepts understood as mental states or structures.

1 People working in the tradition of concepts as mental representations appeal to functional roles or mental shape in order to deal with these problems (Fodor, 1995; Tye and Sainsbury, 2012). For a discussion, see below.

2 Millikan (personal communication) thinks her view should be better classified in the first group, that is, as holding that concepts are mental representations. I will discuss her view in some detail at the end of this chapter.

Secondly, even those who think that concepts should be better identified with abilities (e.g. Kenny, 2010; Millikan, 2000) or Fregean senses (e.g. Peacocke, 1992) admit that there are mental vehicles that carry representational content. So, while they might disagree as to whether these vehicles should be called concepts or not, the naturalization of the representational content carried by these states might still be considered an interesting project (although probably, according to them, incomplete).

Finally, psychologists tend to think of concepts in this way. For instance, Salomon, Medin and Lynch claimed:

> A concept might be very difficult to define. However, in this paper, we will to refer to a concept as a mental representation that is used to meet a variety of cognitive functions (Salomon, Medin and Lynch, 1999, p. 99, quoted in Machery, 2009)

Similarly, Susan Carey (2009, p. 5) wrote:

> I take concepts to be mental representations- indeed, just a subset of the subject's entire stock of a person's mental representations. (...) I assume representations are states of the nervous system, that have content, that refer to concrete or abstract objects (or even fictional entities) to properties and events.

More generally, Machery (2009, p. 10) describes how psychologists tend to think of concepts:

> By 'knowledge', psychologists mean any contentful state that can be used in cognitive processes. (...). Psychologists often characterize concepts as those bodies of knowledge that are stored in long-term memory and that are used in the process underlying the high cognitive competences.

Obviously, assuming the notion of concept used in psychology has clear advantages when trying to address and discuss certain issues from the psychological literature, as I will often do in the remainder.

Summing up, in what follows I will be assuming that concepts are mental states that compose thoughts and other propositional attitudes. They are the vehicles of thought and possess representational content. That will be our starting point.

We will move now to two other central questions in the debate on concepts. The first issue concerns conceptual structure and the second one conceptual content. Before going into details, it is worth mentioning that, as Rey (1985) pointed out long time ago, very often the debate on the structure of concepts conflates metaphysical and epistemological problems. In particular, I think the question on conceptual structure and the question on conceptual content have usually been confused. Here I will spell out the question of conceptual structure and the question of conceptual content as addressing different issues. At various points I will present some arguments in favor of this way of describing the debate. As we will see, one of the main reasons is that if the question of structure is not properly distinguished from the question of content determination, the views of some philosophers like Prinz, Schroeder or Millikan are hard to understand.

5.1.2 *The Structure of Concepts*

Assuming that concepts are mental representations, the following question I would like to concentrate on is whether concepts are structured or unstructured entities. To a first approximation, concepts are structured entities iff in the possessing conditions of every concept there are other concepts involved. A concept's possession conditions specify the set of necessary and sufficient conditions that a subject must comply with in order to be said to possess a concept. If concepts are structured, then, there is a set concepts that a subject must possess in order to possess a concept (see below for some clarifications).[3]

Consequently, the question of the structure of concepts boils down to the following: are concepts constituted by other concepts, such that one cannot possess concept C unless she also possesses a different set of concepts S? How are these constitutive concepts assembled? These and similar questions have received a lot of attention in the last decades. Probably, one reason for the widespread interest on conceptual structure is that this issue is very closely related to the debate on how concepts are learned (Fodor, 1998; Margolis, 1998; Lawrence and Margolis, Forthcoming). How we learn a concept reveals its structure, and its structure determines which process must be followed in order to acquire it.

The view that concepts are unstructured entities is usually called 'Conceptual Atomism'. Fodor and Millikan (see below), for instance, endorse this account. In contrast, the range of views that take concepts to be structured entities adopt different names depending on the kind of structure concepts are supposed to have. The Prototype Theory, the Stereotype Theory or the Theory-Theory are some examples. Since all these approaches share the common assumption that concepts are structured, I will call them 'Conceptual Structuralism'.[4]

Let me define in more detail each of these views.

5.1.2.1 *Conceptual Atomism*

In several books and papers, Fodor (1998, 2004, 2008) has defended a view that has come to be known as 'Conceptual Atomism'. He defines Conceptual Atomism as the view that "satisfying the metaphysically necessary conditions for having one concept never requires satisfying the metaphysically necessary conditions for having any other concept" (Fodor, 1998, p. 14). Conceptual Atomism claims that the possession conditions for a concept are independent from the possession conditions of any other concept; concepts are unstructured entities or *atoms*. According to this approach, two subjects can have the concept DOG

3 For instance, Fodor (2008, p.25) writes: "'Concepts are (or aren't) definitions' is the way the issue is usually framed in cognitive-science literature. Probably, the claim that's actually intended is about concept possession; something like: 'to have a concept is to know its definition'."

4 It will not escape to many readers that Conceptual Structuralism is very close to what some people call 'Conceptual Role Semantics' (Block, 1986). However, Conceptual Role Semantics is usually understood as a view not only on the conceptual structure, but also on conceptual content. Since here I want to keep questions about structure and questions about content separated, I will coin this neologism in order to refer to those accounts that deny Conceptual Atomism.

even if they do not share any other concept, since there is no concept that a subject is required to possess in order to possess DOG.[5]

There is however an important issue in this debate that should be taken into account when formulating Conceptual Atomism and the alternative views. It is a platitude that Conceptual Atomism is false of some concepts like BROWN COW. In order to possess the complex concept BROWN COW one surely needs to possess the concept BROWN and the concept COW (Fodor, 2008, p. 141). Similarly, it is almost trivially true that some concepts are atomic, in the sense that we do not need to possess other concepts in order to possess them; even supporters of the Classical Theories (see below) thought that there must be some sensory concepts (e.g. RED, HOT, ...), which are basic and are used in order to define the rest of concepts. So, if everyone accepts that some concepts are structured and some are not, what is the real disagreement between these theories?

The real discussion is on whether concepts like CAR, WATER, MAMA, DEMOCRACY or TIGER are atomic or not. Since it is almost a platitude that some concepts are atoms and some are not, the debate is on whether everyday concepts are structured. The controversial question is whether this large set of common concepts that we employ in everyday life are atomic. Consequently, for lack of a better word, I will call this set of concepts 'standard' concepts.[6]

So here is a plausible definition of Conceptual Atomism:

First Atomism For any standard concept C and any set of concepts S,[7] a subject can have C without having S.

First Atomism works as a first approximation to Conceptual Atomism, but there is an important ambiguity in the scope of the quantifier that allows for a stronger and a weaker reading. One the one hand, First Atomism could be stating that no concept *whatsoever* is involved in the possessing or individuating conditions of a given concept (Weiskopf, 2009). This reading is compatible with the possibility of the subject having just one concept. The second reading suggests that *no particular*

5 As Fodor (2008, p. 141) suggests: "The metaphysics of concept possession is atomistic. In principle, one might have any concept without having any of the others (except that having a complex concept requires having its constituents concepts)."

6 Some people might worry that the real question is not about *standard* concepts, but about *most* concepts, *simple* concepts or *lexical* concepts. However, while the notion of 'standard' is not ideal, I think it is still better than these alternative proposals. Let me explain.
On the one hand, if the debate were cashed out in terms of 'simple concepts' (something like the following: 'Conceptual Atomism is the view that simple concepts are unstructured'), that claim would be trivially true. After all, everyone accepts that if a concept is simple (i.e. non-composed), then it is unstructured. The real question then, would be whether concepts like WATER are simple or not.
Secondly, spelling the debate in terms of 'most concepts' (something like 'Conceptual atomism is the view that most concepts are unstructured') is tricky, because it seems that Fodor or Millikan (which are conceptual atomists) need not accept that *most* concepts are atoms. For instance, they could hold that, as a matter of fact, most of our concepts are composed. That is, perhaps most of the time we use concepts like BROWN COW or TASTY WATER. I doubt the question between atomism and non-atomism turns around the *number* of concepts that are unstructured. That would lead to a very different kind of debate from the one that has been taking place.
Finally, many people appeal to lexical concepts, that is, concepts that are expressed in English using a simple lexical expression. But, again, that seems to assume a particular view on the relation between thought and language; conceptual atomists need not accept that there is simple concepts in thought are always expressed by lexical expressions in English.
The real debate concerns concepts like BUILDING, PHONE or TREE. So, for need of a better name, I call them 'standard concepts'.

7 Obviously, the set of concepts S cannot include C.

set of concepts are involved in a concept's possessing or individuating conditions. On that interpretation, there is no particular set of concepts I need to possess in order to have the concept DOG, even if I might be required to possess at least a certain amount of concepts.

I think the second reading is preferable for three main reasons. On the one hand, not even Millikan, Margolis or Fodor would accept that no concept whatsoever is involved in the individuating conditions of standard concepts (Fodor, 1975). As a result, adopting this strong reading would result in a view that nobody defends.

A second reason for favoring the weaker reading of First Atomism is that, on a very popular understanding of what concepts are, *by definition* concepts are entities that can be multiply combined with many other concepts that a subject possesses (Evans, 1982). Similarly, Fodor (1986) argued that having an inferential capacity (which surely requires more than one concept) is a constitutive part of having a representational system (which distinguishes our representational capacities from the capacities of paramecia). Hence, a subject cannot have a single concept.

Another reason is that assuming the strong reading would probably preclude a plausible answer to Fregean puzzles. The concern here is that in order to solve Fregean puzzles we need to accept that concepts should not be exclusively individuated by referential content. As I argued earlier, my concept CICERO and my concept TULLY are different concepts but they have the same referents, so the individuating conditions of concepts cannot exclusively involve referential content. If we reject the appeal to senses or narrow content (as conceptual atomists usually do; see Fodor, 1995, 1998, 2008; Millikan, 1993, 2000) a plausible solution to this worry postulates some kind of inferential dispositions or functional role. Since the inferences I am disposed to make with the concept CICERO differ from the inferences I am disposed to make with the concept TULLY, they are different concepts.[8] The point I want to stress is that inferences necessarily involve other concepts, so a plausible solution to Fregean puzzles implies that having other concepts is required in the individuating conditions of concepts (even if no particular set of concepts is needed). Therefore, an answer to the problem of distinguishing coreferential concepts is likely to entail that other concepts are somehow involved in concept individuation.

Note that this functionalist solution to Frege puzzles is compatible with the claim that no *particular set* of concepts figure in a concept's individuation conditions; the mere assumption that the concepts associated with TULLY are very different from the concepts associated with CICERO might suffice for accounting for the fact that they are different concepts. It might still be true that there is no particular set of concepts that one must possess in order to possess TULLY.

Therefore, I suggest to formulate Conceptual Atomism as the claim that there is no particular set of concepts involved in the possession conditions of a standard concept. For example, for a subject to possess the concept BIRD, there is no particular concept she has to possess, not even ANIMAL or EGG-LAYER.

Accordingly, a better definition of Conceptual Atomism is:

Second Atomism For any standard concept C, *there is no particular set of concepts S*, such that a subject needs to posses S in order to possess C.

8 This is not the only solution, however. Sometimes Fodor seems to merely be appealing to the *shape* of the mental words in order to solve Frege puzzles (2008).

Finally, I think this definition needs a last amendment. If, as I argued, certain inferential roles are required for individuating concepts, and assuming that inferences always involve a limited set of logical concepts (AND, OR,...), there may be after all a set of concepts that are necessarily involved in the individuating conditions of any concept, namely the set of logical connectives. Of course, Conceptual Atomism is usually meant as a thesis about non-logical concepts (as the view that I can possess BIRD without possessing ANIMAL or EGG-LAYER), so I think we can reach a much accurate definition of Conceptual Atomism if we exclude logical constants. The final proposal, then, is to formulate Conceptual Atomism as follows:

CONCEPTUAL ATOMISM For any standard concept C, there is no particular set of *non-logical* concepts S, such that a subject needs to posses S in order to possess C.

I think that CONCEPTUAL ATOMISM could be accepted by most conceptual atomists, e.g. Fodor, Margolis and Millikan, and would be rejected by non-atomists (Prinz, 2002; Peacocke, 1992; Weiskpof,2009, 2009; Mandler, 2004). So, I am going to assume that CONCEPTUAL ATOMISM captures the view that concepts are unstructured entities.

There are three main arguments in favor of conceptual atomism. First of all, Conceptual Atomism is in a very good position for explaining the compositionality of thought (Fodor, 1998, 2004). It seems that if I have the concept RED and I have the concept BIRD, I am already in a position to use the concept RED BIRD without further aid. In general, it seems that the only thing that is required in order to have a complex concept is to possess the composing concepts (and knowing how to put them together). If concepts are conceived as unstructured entities, it is very easy to explain how that is possible (we will see that conceptual structuralism has problems satisfying this desideratum).

A related argument in support of Conceptual Atomism is that it fits very well into the usual way of describing concepts as mental words. As I said, the Representational Theory of Mind, which is a popular view in cognitive science, holds that thoughts are structured in a language-like manner, i.e. that there is a Language of Thought (Fodor, 1975, 2008). Now, if thoughts are conceived as having a language-like structure, concepts are naturally described as mental words. The idea that concepts are atoms and the idea that they are mental words naturally reinforce each other. Indeed, a common argument in favor of the Language of Thought Hypothesis is precisely the compositionality of thought, which was the first argument I pointed out in favor of Conceptual Atomism.

Thirdly, prima facie it seems that many people that have the same concept DOG differ very much in the properties they attribute to dogs. So probably, the burden of the proof is on those who think that there is a set of privileged concepts that constitutively determines whether or not a subject possesses DOG.

Of course, Conceptual Atomism is not devoid of problems. However, I think that most objections come from a conflation between the question on conceptual structure with the question of content determination, as I will argue below (see 5.1.2.3).

Probably, the most serious objections against Conceptual Atomism concern certain counterintutive consequences in extreme cases. For instance, according to Conceptual Atomism it is possible to have the

concept BIRD and think it is a piece of furniture. But that sounds odd. Intuitively, it seems that possessing the concept BIRD requires something more than just having a mental word that refers to birds. In particular, it seems I cannot be completely mistaken about the kind of entity I have a concept about. Conceptual Atomism must find a way of addressing this difficulty. In 6.4.4 I will address this issue with the tools presented in chapter 6.

Let us move now to Conceptual Structuralism.

5.1.2.2 *Conceptual Structuralism*

Conceptual Structuralism can be defined as the denial of Conceptual Atomism, so it claims that there is a particular set of concepts that a subject must possess in order to possess a standard concept. In other words:

Conceptual Structuralism For any standard concept C, there is a particular set of non-logical concepts S, such that a subject needs to posses S in order to possess C.

Conceptual Structuralism is the standard view among psychologists (Machery, 2009; Mandler, 2004; Carey, 2009) and some philosophers (Prinz, 2002; Clark and Prinz, 2004; Weiskopf, 2009). Nonetheless it is important to distinguish different versions of Conceptual Structuralism, depending on the set of concepts that are supposed to be part of the possessing conditions of concepts and its relations. Here I will discuss the Classical Theory, the Prototype Theory and the Theory-Theory, which are the most important ones.

Classical Theory

The Classical Theory was first put forward in Plato's Eutyphro and has been the most important approach until the XXth century. Even if nowadays few people endorse this view, I think it is useful to consider it, both because of its historical (and intuitive) support and because it is the theoretical basis on which posterior theories like the Prototype Theory were built.

In a nutshell, the Classical Theory holds that standard concepts have definitions and that in order to posses a concept one has to know this definition. More precisely:

Classical Theory For any standard concept C, there is a particular set of non-logical concepts S, such that a necessary and sufficient condition for a subject to possess C is that it possesses S. All and only members of S are involved in the definition of C.[9]

On the Classical Theory, concepts are individuated by appealing to a set of concepts that specify necessary and sufficient conditions for concept possession. In other words, for any concept C there is a set

9 Lawrence and Margolis (2011) define the Classical View as follows: "According to the classical theory, a lexical concept C has a definitional structure in that it is composed of simpler concepts that express necessary and sufficient conditions for falling under C. (...). The idea is that something falls under BACHELOR if it is an unmarried man and only if it is an unmarried man". So they describe the debate between structuralists and atomists as a discussion between two views on possession conditions *and content*. As I said, I think it is better to keep the debate on structure and the debate on content determination separate.

of non-logical concepts S, such that a subject possesses concept C iff she also possesses S. Thus, a subject has the concept GOLD iff she also possesses the concepts SOLID, YELLOW, METAL, and so on.

The first and intuitive virtue of the Classical Theory is that it can clearly explain concept acquisition and categorization using the same features (Lawrence and Margolis, 2011). When we teach a concept, we typically enumerate a set of features that seem to be definitory of the exemplars falling under this concept. If my child does not know what a tiger is, I will tell him that it is an stripped animal, has four legs, lives in the savannah, etc. The idea of the classical theorist is that I possess the concept TIGER iff I know the defining characteristics of tigers. For this reason, teaching a concept basically consists in communicating the associated description. Having the concept TIGER just is knowing this set of properties that all and only tigers share.

Another central aspect that explains the large life of the Classical Theory despite its obvious difficulties is that it fits perfectly well into the empiricist framework (Armstrong et al., 1983; Lawrence and Margolis, 1999). Empiricists contend that complex ideas are composed of more basic ones, which at the end are reducible to sensory data (Locke, 1690). On this view, most concepts are complex ideas that derive from primitive ones. Thus, many concepts that at first glance seem basic (such as BIRD) are in fact internally structured and decomposable into more basic notions (FLY, SING, FEATHER,...). The empiricist hopes that this decomposition could be carried out until a basic level of purely sensory concepts is reached.[10] So, the Classical Theory on Concepts and Classical Empiricism strongly support each other.

Moreover, there is a further reason in favor of this account. At least since Kant (1787), some people have maintained that there is a kind of judgments called 'analytic' in which the meaning of the predicate is included in the meaning of the subject, e.g 'bachelors are not married'. The Classical Theory offers a way of spelling out the idea of analyticity; the claim that in the sentence 'bachelors are not-married' the predicate is stating something already included in the subject makes sense because the complex concept BACHELOR can be decomposed into the two more basic concepts NON-MARRIED and MAN. Consequently, the Classical theory can explain quite straightforwardly central cases that elicit the intuition of analyticity.[11]

Last but not least, the Classical Theory helps to justify the method of Conceptual Analysis (Jackson, 1998; Bealer, 1998). Paradigmatic conceptual analysis offers definitions of concepts that are to be tested against potential counterexamples that are identified via thought experiments. To the extent that this task can be successfully carried out, this project conveys support for the theory.

However, despite these advantages and the fact that the Classical View has been the predominant view in the history of philosophy, nowadays it is widely discredited because of the insurmountable difficulties it faces. Some of these problems are the following:

10 An interesting consequence of this view is that there must be a set of basic concepts that lack definitions. This is one of the reasons the empiricist picture is committed to the idea that at least some concepts are atomic.

11 Nevertheless, Frege (1884) convincingly showed that there are other analytic truths that cannot be explained with the *containment* paradigm. The analytic truths expressed by sentences like 'Anyone who's an ancestor of an ancestor of Bob is an ancestor of Bob' or 'If something is red, then it's colored' cannot be accounted for by merely adopting Classical Theory.

LACK OF DEFINITIONS After around two millennia of thinkers endorsing the Classical View, hardly any definition of any standard concept has been provided (Smith and Medin, 1981; Rey, 1985, p. 239; Fodor, 1998). Famous proposals have been refuted or are highly controversial: GOOD, KNOWLEDGE, HUMAN BEING or SPECIES are some examples. Of course, *per se* the fact that definitions have not been found does not mean that they do not exist. Nevertheless, it should at least raise some doubts on the veracity of the Classical View.

ANALYTICITY One of the main appeals of the Classical Theory, its explanation of analyticity, has been seriously threatened by Quine's attack on this notion. On his highly influential paper called 'Two dogmas of Empiricism' (1953), Quine undermined the distinction between analytic and synthetic statements. If, as he argued, there is no clear distinction between analytic and synthetic predicaments, the alleged virtue of the Classical Theory becomes a false prediction of the theory. As a result, instead of providing further support for it, it should be regarded as an unwelcome consequence of the Classical Theory (Fodor, 2004).

PROTOTYPICALITY The Classical Theory is unable to explain the prototypicality effects that were discovered and studied in the 1970s. Psychologists like Rosch (1978) pointed out that exemplars that fall under the very same category possess different degrees of exemplariness. For example, people are able to rank exemplars of fruits with respect to how 'good they are' or 'how typical they are', so that an apple seems to be a better exemplar of fruit than a strawberry or an olive. This result is at odds with the Classical Theory, because if an exemplar fulfills the alleged set of necessary and sufficient conditions, it should immediately fall under the relevant category. If being a fruit amounts to having features X and Y and both apples and olives are X and Y, why do we think that apples are better exemplars of fruit than olives? The Classical Theory is unable to explain how there can be degrees of typicality among different exemplars.[12]

These are some of the compelling reasons that have led people to relinquish the Classical Theory. And, nonetheless, instead of completely abandoning the view and embrace Conceptual Atomism, some philosophers and psychologists tried to modify the Classical Theory so as to yield a new account of concepts that could deal with these drawbacks and at the same time keep the idea that concepts are structured. This was the origin of the rest of structuralist theories. Let me discuss this set of Non-Classical Theories.

12 Here is one reason for distinguishing the debate on structure from the debate on content: if, as many people suggest, the discussion on the structure of concepts were a discussion on the specification of the extension of a concept, the objection of protypicality would fail. To see why take, for instance, Earl (2007)'s formulation of the Classical Theory: 'the classical theory of concepts holds that complex concepts have classical analysis, where such an analysis is a proposition that gives a set of individually necessary and jointly sufficient conditions *for being in the possible-worlds extension of the concept being analyzed*' (emphasis added). If the Classical View endorsed that claim, it would be perfectly compatible with prototypicality effects discovered by psychologists; after all, one can coherently hold that there are necessary and sufficient conditions for falling under the extension of a concept (e.g. x is water iff x is H_2O), and at the same time assume that people use prototypes in order to categorize entities. Thus, the standard objection of protoypicality only makes sense if the debate is cashed out in terms of possession conditions.

Non-Classical Theories

Non-Classical Theories (NCT) mainly emerged as a way to deal with the problem of analyticity and protoypicality of concepts. This set of views (developed in different ways in diverse theories) holds that concepts are individuated by a fuzzy set of different concepts (or beliefs), such that possession of a sufficient number of them is necessary and sufficient for possessing the standard concept. Thus, a general definition of the Non-classical Theories is the following:

NON-CLASSICAL THEORY For any standard concept C, there is a particular set of non-logical concepts (or beliefs) S, such that possessing *a sufficient number* of concepts (or beliefs) of S is a necessary and sufficient condition for a subject to possess C.

In other words, the main difference between classical and non-classical theories, is that on the latter the set of concepts that a subject has to possess in order to possess a given C is fuzzy (Osherson and Smith, 1981,p. 35).[13] Some Non-Classical Theories are the Prototype Theory, the Stereotype Theory, the Exemplar Theory or the Theory-Theory.

I think it is not unfair to say that the paradigmatic Non-Classical Theory is the Prototype Theory,[14] which is based on the the notion of a prototype. A prototype is a set of concepts that specify a set of features that a given entity falling under the concept tends to possess (Rosch and Mervis, 1975; Rosch, 1978). For instance, the prototype of CHAIR is constituted by FURNITURE, WOODEN, FOUR-LEGGED, and so on. According to the Prototype Theory, concepts are individuated by a set of properties that entities falling under them tend to possess. One has the concept CHAIR iff one possesses a sufficient number of concepts encoding properties that chairs tend to have.

There are further claims that usually accompany the theory. First, some properties associated with a given concept are *more typical* than others. Consequently, entities can be ranked according to the amount of prototypicality they display (Rosch, 1978). For example, red apples are better instances of APPLE than brown ones, because RED is better ranked than BROWN in the prototype of APPLE (or, alternatively, BROWN is not in the prototype of APPLE). This idea is supported by a large set of empirical evidence. For instance, a classical experiment showed that people are able to rank exemplars of a concept in respect to the typicality they possess and there is a great amount of coincidence among them (Rosch, 1978). Another revealed that there is a slight temporal delay when people have to classify non-typical instances in front of typical ones. That is, experiments show that people tend to classify earlier red apples than brown apples as entities to which APPLE applies. So the Prototype Theory and the other Non-Classical Theories can explain the prototypicality effects exhibited by many concepts.

Secondly, NON-CLASSICAL THEORY assumes that the particular set of features attached to a given concept does not constitute a necessary

13 While a set can be defined as a function from objects to 0 or 1 (depending on whether these objects belong or do not belong to the set) a fuzzy set can be defined as a function from objects to all real numbers between 0 and 1 (depending on the degree to which every object belongs to the set).

14 Some people use 'Prototype Theory' in order to refer to all kinds of Non-classical theories (e.g. Rosch, 1978), while others reserve this name for a particular hypothesis about the mechanisms employed in categorization (e.g. Lawrence and Margolis, 1999; Earl, 2007; Prinz, 2002; Machery, 2009). I will follow the latter convention wich has become standard.

and sufficient condition for its possession. And, since there is no determinate set of concepts one must possess in order to possess a given concept, the strength of finding a definition for all concepts is relaxed. Every concept is associated with a set of features forming a family resemblance (in roughly the sense of Wittgenstein, 1953). So there is no particular concept that must be shared by all concept-possessors (although all subjects must possess a sufficient number of concepts of a certain set).

As I said, there are different theories that would satisfy the description stated in Non-Classical Theory. Prototype Theories, Stereotype Theories, Exemplar Theories, Theory-Theories and many other differ as to how the relevant fuzzy set is determined, what kind of properties are associated with any concept or how it is assessed whether a given mental state satisfies a sufficient number of conditions required for it to qualify as a concept C (for reviews, see Prinz, 2002; Lawrence and Margolis, 1999, 2011; Machery, 2009, Murphy, 2002). Discussing all these approaches in detail would lead us far away from our main purpose here. Nevertheless, I would like to put forward some objections that are probably shared by all of them (with slight modifications).

objections First of all, the treatment of categorization associated with Non-Classical Theories works best for quick and unreflective statements. But when it comes to reflective judgments, people tend to rely much less on similarity, prototypes or exemplars for comparison. For instance, if we are asked whether a dog surgically altered like a racoon is a dog or a racoon, most of us (even young children) would claim it remains a dog (Keil, 1989; Gelman,2003). Similarly with numbers (Armstrong et al. 1983).[15]

Secondly, some experiments show that prototypicality effects are not only found in concepts like APPLE or fruit, but also in well-defined concepts such as EVEN NUMBER, FEMALE or PLANE GEOMETRY FIGURE (Armstrong et. al., 1983). That might show that prototypicality effects have nothing to do with concept possession (or with the determination of the entities falling under them).

Thirdly, there are some concepts that seem to lack prototypes, exemplars, stereotypes or theories. For instance, THING, NOT A WOLF, A CONSEQUENCE OF A PHYSICAL PROCESS STILL GOING ON IN THE UNIVERSE or FROG OR LAMP (Fodor, 1998, pp.101-2; Earl, 2007; Robbins, 2005 p. 271). If it is granted that a subject can possess these concepts even if they lack prototypes, exemplars, stereotypes or theories, then none of these features seem to be required for a subject to possess concepts. So the Non-Classical Theory is in trouble.

Fourthly, in order for a subject to possess a complex concept (such as WHITE COW) it seems to suffice that she has the composing concepts (WHITE and COW) and knows how to put them together. But prototypes, stereotypes, theories or exemplars do not seem to compose in this way (Fodor, 1998,p. 55, p. 102-8; 2001, 2004; Osherson, 1981; Rey, 1983; but see Prinz, 2002, 2008). Suppose I have a concept A, whose prototype is P and a concept B whose prototype is R; since I posses concept A and B it is very plausible that I can form a complex concept AB. However, in some cases, it might happen that the prototype of AB is, say, PQ and not, as we would expect, PR. If possessing the

15 The Theory-Theory, which claims that the possession condition of concepts includes something like theories, can partially avoid this objection.

concepts that determine the prototype where required for possessing the concept, having A and B would not suffice for having AB. However, that is in tension with the claim that all I need in order to possess AB is to possess concept A and concept B.

Fodor illustrates this point with the example of PET FISH. The prototype of PET is something like a dog or a cat, the prototype of FISH is a grey and slender animal (like a trout), but the prototype of PET FISH is a colorful animal (like a goldfish). If prototypes determined the possession conditions of concepts, having PET and FISH would not suffice for having the concept PET FISH, what seems extremely implausible. Other examples are MALE NURSE and RED HAIR (Fodor, 2001).[16]

A final objection to Non-Classical Theories is that of *holism* (Fodor and Lepore, 1992). The problem of holism can be put as a dilemma: either any alteration in the conceptual network (or theories) implies a change in the concept or there is a core set of concepts (or theories) that determines concept possession (and concept identity), which remains identical most of the time (Landau, 1982). If the Non-Classical theorist takes the first horn of the dilemma, interpersonal and intrapersonal identity of concepts becomes almost impossible, because our conceptual networks continuously suffer changes and additions. Alternatively, the second horn consists in holding that there is a conceptual core, which specifies the information that determines concept possession and individuates the concept. Prima facie, that might look like a sensible option, since this set could account for the compositionality of concepts, could explain which entities fall under the concept and so on (e.g. Osherson, 1981). The problem, of course, is that taking this horn just raises the same questions at a different level: what is the structure of this conceptual core? Either if we adopt the classical view or some kind of non-classical view, the same range of problems arise (Lawrence and Margolis, 2011). Furthermore, a criterion should be given as to how this small set of privileged concepts is determined. It is not clear whether Non-Classical Theories can provide a satisfactory criterion.

There is still a vivid debate between Conceptual Atomists and different versions of Conceptual Structuralism. So far, I have reviewed the main views on the matter. Let me draw some conclusions that I think are important for our discussion.

5.1.2.3 *Conclusions on the structure of concepts*

First of all, it is important to notice that both Conceptual Atomism and Conceptual Structuralism can accept that possessing a certain concept (e.g. BIRD) is usually correlated with the possession of other concepts (e.g. FEATHER, FLY). Furthermore, both are compatible with this set of concepts exhibiting different degrees of prototypicality (Millikan, 1998, p . 91). This fact contrasts with popular arguments

16 Fodor (1998) suggests a similar argument in terms of the compositionality of (referential) content, but as Clark and Prinz (2004) argue, Non-Classical Theories have a satisfactory solution. The reply on behalf of Non-Classical Theories is quite simple indeed: since the debate is on the possession conditions for having a concept, these theories can accept that referential content composes. They can coherently hold that the referential content of PET FISH derives from the content of PET and the content of FISH, and nevertheless the possession conditions for PET FISH involve its prototype.
This is another reason for distinguishing the debate on conceptual *structure* from the debate on conceptual *content*.

against conceptual atomism that can be easily found in the literature (Weiskopf, 2009). For instance, Schröder (1998, p. 84) argues:

> According to prototype theory it is the matching process between features of the exemplar and features of the concept that is crucial for the explanation of typicality. Not only is there nothing in Millikan's account that could be analogous to this process, but even if there were it would be unclear on what it would operate: it cannot operate on features, because there are supposed to be none.

And Prinz, (2002, p. 99):

> The greatest shortcoming of atomism involves categorization. Unstructured mental representations simply cannot explain how we categorize. First, consider category production. If concepts are structured, the properties named in describing a category can be associated with features contained in the concept representing that category. If concepts have no constituent features, our ability to produce such descriptions must be explained by information that is not contained in our concepts. Now consider category identification. On most views, our ability to identify the category of an object depends, again, on features contained in the concept for that category. (...). The atomist says that an explanation of categorization is not within the explanatory jurisdiction of a theory of concepts.

However, Conceptual Atomism need not deny that, as matter of fact, concepts are very often (or even always) connected to other representations in that way; rather, the point is that none of these connections is essential for possessing the concept. Conceptual Atomism is also compatible with the psychological data concerning the existence of prototypes. The fact that people usually use prototypes or exemplars in order to cateogize entities does nothing to show that these prototypes are necessary for the subject to possess a given concept. Fodor or Millikan can easily accept that protoypes exist (indeed, they do Millikan, 2000, ch. 3; Fodor, 2008). In contrast to what Prinz says, if 'theory of concepts' is broadly understood as a theory of what concepts are and how they relate to the rest of our cognitive abilities, Conceptual Atomism can also accommodate the capacity for categorization within a theory of concepts. There is no reason for thinking that, in general, a theory of an entity can only include its essential properties. One's conception of cars includes many other properties of cars, beside the essential ones (Weinberg, 2003, p. 282).

The interesting debate is on whether for any standard concept there is a set of concepts that determines a (fuzzy) set of necessary and sufficient conditions for concept possession. It is a disagreement about modal claims, not about what concepts people tend to have. That means that, in contrast to standard ways of presenting the debate (Prinz, 2002 Machery, 2009), Conceptual Atomism is not ill-suited for accounting for the psychological data.

The second important observation is that the main problems of Conceptual Atomism and Non-classical Theory seem to point at opposite directions. On the one hand, Conceptual Atomism puts no limits on the kind of concepts one has to possess in order to possess

a given concept C, and hence it yields counterintuitive results. For instance, it seems to imply that someone might have the concept TREE and think it is a kind of star. Prima facie, something more seems to be required for a subject to have a concept. But, on the other hand, neither definitions nor prototypes, exemplars, theories or stereotypes seem to be the kind of conditions that can reasonably be required, since it is possible to think of many cases where a subject has the concept but does not possess any of these other concepts.

Now, I think that this tension can be partially resolved in favor of the atomist. The chief idea I will defend is that, even though the conceptual atomist has to bite de bullet and admit that it is metaphysically possible for a subject to have the concept BIRD and think birds are a piece of furniture, several mechanisms explain why this case is not as threatening as it might seem. A presentation and partial defense of this view will be offered and defended in the next chapter (see 6.4.4).

The more pressing question for us, though, is whether any of these views on conceptual structure bear on the question on content determination. But, before presenting the interesting relations between theories of conceptual structure and theories of conceptual content, let me define in more detail the different views on conceptual content.

5.1.3 *The content of concepts*

The third topic I would like to address is: What determines the content of our concepts? There are two main positions in this debate. First of all, what I will call 'Semantic Atomism' holds that the content of a concept is not even partially determined by its relation to other concepts. In contrast, what I call 'Semantic Descriptivism', holds that conceptual content is also determined by a set of concepts that accompany a given concept. Let us define each view in more detail.

5.1.3.1 *Semantic Atomism*

Semantic Atomism is the claim that the *content* of a concept is not determined by its relation to any other concept, but (usually) by some informational, functional, causal or covariance relation between the concept and its referent.[17] In short, Semantic Atomism can be spelled out in the following way:

First Semantic Atomism The content of a concept is not determined by its relation to other concepts.[18]

Semantic Atomism is usually conflated with Conceptual Atomism (for instance, Lawrence and Margolis, 1999, 2011; Schneider, 2011, p. 159-81), but I think it is important to keep them distinct. Conceptual Atomism

17 Consider, for instance, Fodor's (2008, p. 54) words: "By contrast, it is plausible prima facie that reference is atomistic; whether the expression 'a' refers to the individual a is prima facie independent of the reference of any other symbol to any other individual.".

18 Again, contrary to my usage here, Lawrence and Margolis (2011) call this view 'Conceptual Atomism', so they do not distinguish the question about structure from the question about content. I think that my definition captures much better the way in which these notions are usually used in the debate (Fodor, 1998; Prinz, 2002; Weiskopf, 2009). Furthermore, if we follow Lawrence and Margolis' usage, we will not be able to understand theories like Prinz (2002), who denies Conceptual Atomism but wants to stick at (a weak version of) Semantic Atomism (for a discussion, see 5.2.2).

is a thesis about the relevance of other concepts in concept possession, while Semantic Atomism is a thesis about content determination.[19]

Now, some contenders in this debate think that all content is referential content (Fodor, 1998, 2008; Millikan, 2000), while others distinguish referential content from *cognitive content* (Prinz, 2002; Weiskopf, 2008). A concept's cognitive content is fully determined by the links a concept has with other concepts in a network. Two people have the same cognitive content if they have identical prototypes. For instance, the cognitive content of the concept WATER is constituted by the concepts COLORLESS, TASTELESS,.... Cognitive content is supposed to be what doppelgängers in twin-Earth cases have in common.

Now, one cannot be a Semantic Atomist about cognitive content, since by definition this content is determined by the relations a concept has in a network. Ex hypothesi, cognitive content is determined by the set of concepts one possesses at a given time, so Semantic Atomism would be trivially false if it was interpreted as a claim about cognitive content. Consequently, in order to make sense of the debate on Semantic Atomism, we need to assume that it is a claim about referential content, not about cognitive content. Furthermore, in this dissertation we have been focusing on the naturalization of referential content, so this is the aspect of conceptual content that should concern us. Semantic Atomism, then, will be cashed out as the claim that a concept's *referential* content is not determined by any relation to other concepts:

Semantic Atomism The *referential* content of a concept is not determined by its relation to other concepts.

The denial of Semantic Atomism is what I will call 'Semantic descriptivism'.

5.1.3.2 *Semantic Descriptivism*

Semantic Descriptivism claims that other concepts play an important role in fixing conceptual content. In that respect, we should distinguish a strong from a weak form of descriptivism:

Strong Semantic Descriptivism The referential content of a concept is *fully* determined by its relation to other concepts.

Weak Semantic Descriptivism The referential content of a concept is *partially* determined by its relation to other concepts.

Before discussing the plausibility of any of these views, let me clarify the relations between these theories of content determination and the view on the structure of concepts.

19 Here is a random example from Weiskopf: "[Atomistic theories] claim that concept possession is not based on the inferences one draws with a concept, but rather with what the concept picks out in the world. Concepts for atomists are fundamentally a kind of category detector. (...) Because these detectors can reliably inform a creature about the world around it, this approach is sometimes termed 'informational semantics', and so atomism may be thought of as an information-based rather than an inferentialist approach. Informational views and inferentialist views differ on whether the fundamental role of concepts is to detect categories in the environment or to facilitate concepts concerning categories. The ability to reliably detect a category does not presuppose the possession of any nother concepts in particular, so atomists do not need to posit the existence of conceptually necessary conditions" (Weiskopf, p. 6)
In the same passage, Wieskopf defines concepts in terms of possession conditions, content determination and in terms of the fundamental role that concepts play.

Many people have assumed that Conceptual Atomism and Semantic Atomism go hand in hand, while the same is true of Conceptual Structuralism and Semantic Descriptivism. I think this is the main reason why people have tended to link theories about conceptual structure with theories about conceptual content. But is there an a priori link between them? Does Semantic Atomism or (any version of) Semantic Descriptivism entail any particular view on the debate on the structure of concepts?

Let us carefully consider all options. First of all, does the acceptance of Semantic Atomism commit one to Conceptual Atomism or to Conceptual Structuralism? That is, if one holds that referential content is not determined by its connection to other concepts, must one accept or deny that there is a set of concepts involved in the possession conditions of any concept? Prima facie, it seems that accepting Semantic Atomism does not commit one to any of these views. On the one hand, Semantic Atomism is obviously compatible with Conceptual Atomism, that is, one could coherently claim that no set of of concepts figure in the possession conditions of a concept and also that no other concept plays a role in the determination of content. Fodor (1998), for instance, endorses this view. On the other hand, it seems perfectly coherent to hold that referential content is determined by a causal or nomological relation between concepts and their referents, and nevertheless possession conditions depend on other concepts. Prinz (2002)[20] and Schneider (2011), for instance, seem to hold such a view, which the latter calls 'Pragmatic Atomism'. As Prinz (2002, p. 257) claims 'one does not need to be an atomist to be an informational semanticist'. So Semantic Atomism is compatible with any view on the structure of concepts.

Let us ask the second question. Does (any version of) Semantic Descriptivism entail any view on the structure of concepts? I think it clearly does not. On the one hand, Semantic Descriptivism is clearly compatible with Conceptual Structuralism. If one holds that in order to possess a concept C one must also possess a set of concepts S, one can coherently hold that this set of concepts S helps to determine content. Indeed, that was the traditional view on concepts until the XXth century. On the other, Semantic Descriptivism is also compatible with Conceptual Atomism. One could hold that there is no particular set of non-logical concepts that one needs to possess in order to possess a given concept, and nevertheless claim that the concepts that one happens to have plays a role in determining a certain concept's content. I will argue that probably Millikan and Papineau endorse this sort of view (see 5.2.4.3).

Therefore, the debate on conceptual content and the debate on conceptual structure are in principle independent debates, in contrast to what some people suggest (Lawrence and Margolis, 2011; Earl 2007). While standard introductions to the debate claim that there are only two options available on the debate on conceptual structure and content (atomism and structuralism/descriptivism) I hope I have convincingly shown that there are at least four views that one could possibly hold, that result from the combination of Conceptual Atomism and Conceptual Structuralism on the one hand, and Semantic Atomism and

20 As he (Prinz, 2002, p.123) claims: 'Perhaps we can accommodate all of the desiderata if we combine the informational component of informational atomism with a nonatomistic theory of conceptual structure.'

Semantic Descriptivism on the other. The usual conflation between debates on the possession conditions of concepts and the conditions that determine content has obscured this fact. Of course, given that one adopts a certain view on one debate, certain possibilities in the other become more plausible. But even if not all views are equally plausible, it is a remarkable fact that many combinations are theoretically possible.

Since this thesis is about the naturalization of content, I will focus my attention on theories of content determination. Nonetheless, I will usually classify the views that I will discuss using these categories, and we will see that they have a crucial importance in some of the arguments that follow.

In the remainder of this chapter, I will present some naturalistic views of conceptual (referential) content and I will show why all of them fail. In the next chapter, I will defend my own view on that matter.

5.2 Towards a theory of conceptual content

After having surveyed the main theoretical options on the nature, structure and content of concepts, let us review some of the most popular naturalistic accounts of how the content of concepts is determined. All of these views assume that concepts are mental representations, but some of them also have strong views on the structure of concepts, which influence some of their arguments. I will make these assumptions explicit when they are relevant.

5.2.1 *Theories of Conceptual Content and Psychology*

The question about how conceptual content is determined is a central question not only in philosophy, but also in certain scientific enterprises. Many scientists are indeed compelled to address this issue, even if most treatments of fundamental aspects of representational phenomena are less than satisfactory. Let me illustrate this statement with some examples.

Since most experimental studies in developmental psychology use 2-D objects in screens in order to investigate the cognitive capacities of infants, some scientists wonder whether we should describe the infant's conceptual capacities as being about 2-D objects or 3-D objects (see Pylyshyn, 2003, p.127). For instance, Carey (2009, p.99) wrote:

> Even if Fodor's analysis has problems it works well enough for the purposes of this book. I accept the causal theory of content determination for representations in core cognition, so Fodor's analysis applies to the case at hand. That 2-D individuals cause object-files to be activated is dependent on the causal relations that ensure that object-files refer to 3-D objects; and in the case of core cognition (unlike concepts such as *cow*) we have at least a sketch of what the relevant causal processes are. Through natural selection input analyzers have evolved that create representations of objects from the information in the physical stimulation of sense organs. It's clear how Fodor's asymmetric dependence theory allows that 2-D entities be misrepresented as objects, and there is evidence it is on the right track.

Carey *claims* that she accepts Fodor's asymmetric dependence theory, but when she spells out in more detail in virtue of what process the asymmetric dependence relation holds, she appeals to the fact that natural selection has selected these input analyzers in order to create representations of objects. Obviously, the latter looks very much like a teleosemantic account (indeed, it seems to be tentatively pointing at the approach defended in part I of this thesis). So it is not clear which of the two approaches she is actually relying on (at that stage of the thesis it should be apparent that the two theories are not equivalent).

Furthermore, notice that Carey's account of why infants represent 3-D objects is not very explanatory. She argues that infants represent 3-D objects because natural selection has selected a mechanism for the production of 3-D object representations; but, as we know, one might reasonably ask why should we think natural selection has endowed us with a mechanism for representing 3-D objects rather than 2-D objects. This is very close to the question we started with. In other words, since it is almost a platitude that perceptual mechanisms have been shaped by natural selection, in order to provide a substantive explanation, she should provide some grounds for us to think that natural selection has created input analyzers that produce representations of 3-D objects, rather than representations of 2-D surfaces (which, obviously, may correlate with the presence of 3-D objects). As we extensively saw in 2.3.3, the claim that a mechanism has been selected for producing a representation of one thing or another requires extensive argument (remember Fodor's black-moving-shadows detectors).

Of course, I am not saying that Carey herself must provide all the details of a naturalistic theory of content. Arguably, she was primary aiming at a psychological theory of conceptual development, so for her theory to work she does not need to get into details of a theory of content (and indeed shes does not attempt to do it). I am just mentioning this example because it illustrates well the fact that the question of conceptual content (1) is of interest to scientists and (2) has been unsatisfactorily addressed in the scientific literature.

In a parallel fashion, Eliasmith (2000, p.7) describes the relation between neuroscience and a theory of meaning in the following way:

> Some of these questions are never explicitly posed by a given school, but all [neuroscience, Fodor, Locke, Descartes and the Stoics] either assume, assert or argue for an answer, and all are appropriate questions to ask. (...) One question these positions definitely share is the problem of how we, qua neurobiological system, have representational content.

Of course, neuroscientists or psychologists are not primarily interested in developing a theory of mental content, so their superficial treatment of the question is partially excusable. But they would welcome very much such a theory. Let us then now consider some explicit attempts to provide a naturalistic account of conceptual content.

5.2.2 *Incipient Causes*

As we saw in the first part of the dissertation, many naturalistic approaches to conceptual content are lead by two intuitions: (1) the idea that some kind of nomic (law-like) relation between concepts and their referent is required for determining conceptual content and (2) the in-

tuition that the causal origin of a concept plays a crucial role in content determination. I think both ideas are mistaken, but it is important to consider why these ideas cannot be made to work. We saw in the first chapter that 1 fails to determine an appropriate content for many representations. Let us discuss now whether 2 or, more interestingly, endorsing 1 along with 2 can help to solve the problems of previous theories when applied to conceptual development.

5.2.2.1 *Prinz (2002)*

First of all, I would like to discuss Prinz's[21] naturalistic account of conceptual content.[22] In the classification I defended previously, he holds CONCEPTUAL STRUCTURALISM and SEMANTIC ATOMISM, that is, he thinks that in order to posses a concept there is a (fuzzy) set of other concepts I need to possess, and, at the same time, he holds that (referential) content is determined by some causal/covariational relation to the world.[23]

As he admits, Prinz's (2000, 2002, 2006) account is intended to be a combination of Fodor's (1990) Asymmetric Dependence Theory and Dretske's (1981, 1986) Informational Theory. According to him, for a concept C to have X as its content (that is, for C to mean X) two conditions need to be met: (1) X must be C's *incipient cause* and (2) there has to be a *nomological covariance* between C and X. Let us define both notions in detail.

On the one hand, Prinz shares the intuition that the naturalization of content should appeal to some kind of causal relation. However, we know that not any causal relation between an entity and a concept will do (remember the problems of misrepresentation and indeterminacy described at length in chapter 1 and 2). Prinz suggests that the relevant cause must be the *first* one. Drawing on etiological theories of direct reference (Kripke, 1980) and inspired by Dretske (1981)'s appeal to a learning period, he thinks that the entity that causally originated the concept is specially important in determining reference. That is why his first condition for content determination appeals to what he calls the 'incipient cause': X is the incipient cause of the concept C iff X caused the formation of concept C. That is:

INCIPIENT CAUSE X is the incipient cause of C iff X is the first cause of C (i.e., X originated the creation of C)

According to Prinz, a necessary condition for C to mean X is that X has been the originating cause of the concept.

Surely, the mere appeal to the incipient cause is insufficient for providing an adequate account of content (we will see that one of the main reasons has to do with problems of indeterminacy). For this reason,

21 Prinz (personal communication) has recently changed his mind at that point. He seems no longer to believe that a naturalistic theory of content can succeed.

22 Let me point out that this naturalistic theory of content has not been much discussed in the literature, even if some of Prinz's main arguments heavily rely on it. For instance, when Prinz (2006) argues that we can perceive abstract entities, he supports his argument with a particular view of how conceptual content is determined. He has also employed this account in his theory on emotions (Prinz, 2004). I would like to show that his theory of content determination falls prey to important difficulties. Again, here I will focus on Prinz's account of *referential* content (that is, truth-conditions), which Prinz distinguishes from something he calls 'Nominal Content' (Prinz, 2000) or 'Cognitive Content' (Prinz, 2002).

23 Again, notice that this approach only makes sense if one distinguishes the question of concept possession from the question of content determination.

Prinz resorts to the tradition that postulates a causal covariance between a concept and its referent. The intuition that the reference relation is determined by some sort of covariance is common and, as we extensively saw, has lead to a range of different proposals (e.g. Dretske, 1981, 1986; Rupert, 2008). However, Prinz's notion of *nomological covariance* differs from other proposals in not being based on a covariance within the actual world, but across possible worlds. According to Prinz (2002, p. 241), nomological covariance has to do with covariance in proximate worlds:

> NOMOLOGICAL COVARIATION Xs *nomologically covary* with concept C when Xs cause tokens of C in all proximate possible worlds where one possesses that concept.[24]

That is, John's concept DOG means *dog* partially because in all proximate possible worlds where John has DOG, tokens of this concept have been caused by dogs.[25]

While NOMOLOGICAL COVARIATION connects with the tradition that seeks to naturalize content by appealing to a covariation between representations and their referents, notice that this particular notion is different from the views we discussed in chapter 1, that is, WEAK INDICATION, STRONG INDICATION, RELATIVE INDICATION and ASYMMETRY. First, according to WEAK and STRONG INDICATION whether C actually covaries with X depends on the number of occasions in which C and X have been coinstantiated. In contrast, NOMOLOGICAL COVARIATION is spelled out in counterfactual terms, and hence it is irrelevant how often X has correlated with C in the actual world.

Similarly, note that unlike RELATIVE INDICATION and ASYMMETRY, NOMOLOGICAL COVARIATION does not take into consideration other possible causes of C. Whether C actually covaries with X only depends on the relation that holds between C and X in nearby possible worlds. Other possible or actual causes of C are not taken into account. Indeed, Prinz's reasons for departing from Fodor's view are very similar to the ones we pointed out in chapter 1 (see 1.2.4).

It should also be clear that NOMOLOGICAL COVARIATION alone is too weak a relation for grounding semantic relations because there are many things mental states nomologically covary with. If in proximate worlds the transparent and colorless liquid that fills oceans and ponds is XYZ, then my concept WATER nomologically covaries with water (H_2O), but it also nomologically covaries with XYZ. More generally,

24 Let me mention that Prinz sometimes adds a 'ceteris paribus' condition, so that he sometimes defines nomological covariance in the following way: 'Xs nomologically covary with concept C when, *ceteris paribus*, Xs cause tokens of C in all proximate possible worlds where one possesses that concept'. I have removed this clause because (as I argued at length in part I), any appeal to *normal conditions* or *ceteris paribus conditions* threatens to undermine the naturalistic credentials of the theory.

25 In Prinz (2002, ch. 9) he suggests that a general motivation for counterfactual theories is to solve the 'Swampman problem'. As we saw in 3.3.4, any theory of content that requires a causal relation in the actual world between X and C in order for C to represent X is committed to denying that Swampman has representational states, because nothing has caused the Swampman's brain states. But Swampman is not a problem for counterfactual theories. While Swampman lacks causal history, it seems his brain states support the same counterfactuals as we do, since *ex hypothesi*, Swampman is microphysically identical to normal humans and the truth of many counterfactuals seem to be grounded on internal properties of human beings. So, in principle, a notion of covariation seem to allow us to attribute representational states (and concepts) to swampbeings.
Unfortunately, Prinz cannot use the solution to the Swampman problem as a motivation for his appeal to nomological covariance, because for him a necessary condition for a concept to mean X is that X is C's incipient cause, and surely Swampman's brain states lack this sort of relation.

anything that sufficiently resembles WATER in proximate worlds would be included in the content of John's concept WATER (that is, John's concept in the actual world would mean *water or XYZ*). That would make concepts highly disjunctive.

So Prinz (2002, p.251) puts together the two notions (incipient cause and nomological covariance), which are supposed to provide necessary and sufficient conditions[26] for content determination:

INCIPIENT THEORY X is the intentional (referential) content of C iff:

1. An X was the incipient cause of C, in accordance with INCIPIENT CAUSE
2. Xs nomologically covary with tokens of C, in accordance with NOMOLOGICAL COVARIATION

There are two nice points in favor of this account. First of all, it seems to yield the right results in a wide range of cases. Take the concept TREE. On the one hand, we might reasonably suppose that we developed this concept when we were confronted with a tree, rather than by seeing a cat or Obama. On the other, it seems that in all proximate worlds where I have this concept, trees still cause it. For instance, if we consider nearby worlds in which trees are a bit higher, or have a different color, or even worlds in which my visual apparatus is slightly different, it seems that trees still cause my concept TREE. Thus, INCIPIENT THEORY gives the right result in many situations.

Secondly, this approach seems to be fully naturalistic. Only causal and counterfactual conditions are mentioned in INCIPIENT THEORY, so there is no intentional notion in the explananda. In that respect, it seems that Prinz's view should not raise any naturalistic qualms.

Let me argue, however, why I think this account is unlikely to be satisfactory.

5.2.2.2 *Discussion*

Let me present four objections against Prinz's view.

INDETERMINACY As we saw in 1.2.2.4, a general way of stating the problem of indeterminacy is the following: a theory suffers from the indeterminacy problem if the theory entails that there are many entities represented by a given state, while common sense and science assume that it has a much more determinate content. Think, for instance, about John's MONARCH concept, that is, the concept that we would naturally attribute to John, which seems to unambiguously refer to monarch butterflies (John uses it when he sees a monarch, and so on). Following Prinz, we can reasonably assume that the incipient cause of John's concept was a monarch and that this concept nomologically covaries with monarchs. However, monarchs are butterflies (indeed, this is a good candidate for being a necessary truth). So if a monarch was the incipient cause of John's concept, so was a butterfly. Thus, if condition 1 is satisfied by a monarch it is also satisfied by a butterfly. Similarly, if in all proximate possible worlds monarchs cause tokens of John's concept, butterflies also do (because monarchs are butterflies). So condition 2 is

26 Let me mention that Prinz thinks that these are necessary and sufficient conditions for the great majority of concepts, but he also claims that it might be the case that other concepts acquire their content in a different way.

also satisfied by butterflies. Therefore, John's concept MONARCH means *monarch or butterfly.*

Indeed, similar results can be obtained with a wide range of properties: insect, animal,... The consequence seems to be that INCIPIENT THEORY entails that the content of John's concept is *monarch or butterfly or insect or....* Indeed, even the property of being a monarch-looking thing causes troubles, since condition 1 and 2 of INCIPIENT THEORY seem also to be satisfied by them: if a monarch was the incipient cause of John's concept, a monarch-looking thing probably was, and if monarchs cause John's mental state in the actual world, monarch-looking things will probably cause John's mental state in close possible worlds. However, this highly indeterminate content starkly contrasts with the original assumption that John had the concept that referred to monarchs (and only monarchs).[27]

Notice that a similar problem will be found in any concept, so the objection generalizes: for any concept, INCIPIENT THEORY entails that it will have a highly indeterminate content. So, even if appealing to incipient causes enables the theory to avoid including entities existing in proximate worlds that resemble very much the entities in the actual world (such as H_2O and XYZ), there are still many sources of indeterminacy that INCIPIENT THEORY cannot exclude.

Interestingly enough, Prinz sometimes seems to be suggesting that, as previously stated, INCIPIENT THEORY can already deal with the serious problems of indeterminacy.[28] However, at other places he adds further conditions in order to deal with this problem.[29] In particular, in Prinz (2002, p. 242-3) he tries to solve what he calls the 'semantic-marker' problem, which basically is a version of the indeterminacy problem suggested earlier. He claims that three further conditions need to be added to INCIPIENT THEORY in order to determine whether a concept refers to a natural kind, an individual or an appearance property (such as *being a monarch-looking thing*):

SEMANTIC MARKERS

(A) C is a kind concept if had Xs looked different than they do, they would still cause tokens of C.

(B) C is an appearance concept if had Xs always looked different than they do, they would not cause tokens of C.

(C) C is an individual concept if were the subject presented with objects that appear exactly like X, at most one of those objects would cause tokens of C. (Prinz, 2002, p. 242-3)

27 Of course, one could say that, in this case, John's concept is not the concept MONARCH, but the concept MONARCH OR BUTTERFLY, etc... If one takes this option, the objection should be better formulated in the following way: Prinz's theory entails that John lacks the concept MONARCH, as well as the concept TREE, WATER, GOLD, and so on.

28 In particular, he writes (Prinz, 2002, p. 241):

> The [second] clause solves the qua and chain problems and can be embellished with further detail about the nature of the nomological relations involved to solve the semantic-marker problem.(...). For example, nomological covariance determines that my MONARCH concept refers to monarchs and monarch mimics but not to butterflies or retinal images, (...).

The qua, chain and semantic-marker problem are different versions of the indeterminacy problem, so in this quote Prinz is claiming that slight refinements in the conditions set up in INCIPIENT THEORY can deal with this problem. Contra Prinz, I have argued that 2 does not solve any of these problems. As we said, not only monarchs covary with C, but also butterflies, monarch-looking things, certain activations in the retina, and so on.

29 Indeed, he presents this proposal as a slight modification of condition 2 in INCIPIENT THEORY. Nevertheless, as we will see, this is in fact a new condition.

If we focus on a) and b), the idea is the following: consider the set of proximate worlds where Xs look different than they look in the actual world. If in these worlds Xs still causes C, then C is a kind concept. If they do not, then C is a concept of an appearance (a concept of X-looking thing). c) tries to apply the same idea to the case of individuals.

The first important thing to notice is that, in contrast to what Prinz claims, a) b) and c) are in fact *new* conditions that should be added to Incipient Theory, rather than embellishments of condition 2. There is an easy way to see why this is so: while condition 2 of Incipient Theory states that we should consider all proximate worlds where a subject still has the concept, clauses a), b) and c) appeal to those worlds where things look a different way, which might be very distant worlds. For instance, if in all proximate worlds Xs still look the same way, in order to assess whether a), b) or c) hold we might have to take into account distant worlds. Nevertheless, in order to see whether condition 2 holds, we should only consider proximate worlds. That shows this solution to the semantic markers problem brings in a new set of clauses into the definition.

Secondly, there is an obvious problem with this view; even if this proposal succeeded, it would provide a recipe for distinguishing concepts about kinds, appearances and individuals, but the problem of indeterminacy is much more widespread. *Monarchs, butterflies* and *insects* are all natural kinds, so merely adding these counterfactual conditions will not rule them out. The conditions set up in Semantic markers are not fine-grained enough for the task at hand. Therefore, Prinz's theory seems to fall prey to the indeterminacy problem, even if semantic markers are added.

Now, whereas I think that Prinz has failed to solve the indeterminacy problem, I would like to explore a possible reply on behalf of Prinz's approach. Basically, the idea is to generalize the strategy of semantic markers suggested in a), b) and c) in order to rule out *any* inadequate properties. The proposal is the following: for any properties X and Y that satisfy conditions 1 and 2 of Incipient theory (that is, for any two properties that cause problems of indeterminacy), consider the most proximate worlds in which one is instantiated but not the other (say, X is instantiated, but not Y). If in those worlds, X still causes tokens of the concept, then C means X (and not Y). If it does not, then C does not mean X. In other words:

Better Semantic Markers For any properties X and Y that satisfy 1 and 2 of Incipient Theory, consider the set of proximate worlds where Xs are not Y.

1. If Xs still cause tokens of C, C represents X (and not Y).
2. If Xs do not cause C, C does not represent X.

That is, in order to know whether John's concept refers to monarchs or butterflies, Better Semantic Markers tells us to consider the possible worlds where there are butterflies but no monarchs; if in those worlds butterflies still cause tokens of John's concept C, then it is a concept of butterfly (and not of monarch); if butterflies do not cause C, then John's concept is not a concept of butterfly.[30] Similarly, in order to

30 Of course, in order to that, concepts should be individuated narrowly.

know whether C is about monarchs or monarch-looking things, look at the most proximal worlds where monarchs are not monarch-looking things;[31] if monarchs still cause C, then C is about monarchs. Otherwise, C is about monarch-loking things.[32]

Now, if we add Better Semantic Markers to the original theory, we get the following account:

Better Incipient Theory X is the intentional (referential) content of C iff:

1. An X was the incipient cause of C, in accordance with Incipient Cause
2. Xs nomologically covary with C, in accordance with nomological covariation
3. For any properties X and Y that satisfy 1 and 2, consider the set of proximate worlds where Xs are not Y:
   a) If Xs still cause tokens of C, C represents X (and not Y).
   b) If Xs do not cause C, C does not represent X. (Better Semantic Markers)

I think this is a better proposal than the previous one, and Better Semantic Markers provide a better reply to the problem of indeterminacy than Prinz' distinction between kind, appearance and individual concepts. Nevertheless, I think that even this refined version of Better Incipient Theory utterly fails to solve the indeterminacy problem. There is an important difficulty that this solution to the semantic marker problem cannot deal with.

First of all, remember why the crude causal account presented in chapter 1 (which claims that a state represents whatever causes it) cannot work: since in the actual world many different entities cause mental states, that would yield a highly indeterminate content. Since my concept DOG is caused by dogs, wolfs and even cats (at dark nights), all these entities would figure in the content of the representations.

Now, the problem of Prinz's theory is that he assumes that by merely moving to other possible worlds, we will be able to distinguish the right cause from the wrong causes; but this is far from clear. Even if we move to other possible worlds, there are many things that cause my concept MONARCH. Some of the things that cause my concept MONARCH are not monarchs, and this is true even if we move to close possible worlds where there are no monarchs. Hence, the problem is the following: even if John's concept meant *monarch,* in some of the possible worlds where butterflies are not monarchs, some of these butterflies cause tokens of John's concept, so condition 3 will not rule out *butterfly* from the content. In other words: if we move to those possible worlds where properties X and Y are not instantiated together, we will probably find out that in some of these worlds X (but not Y) cause C and in some other worlds Y (but not X) cause C. People also make mistakes in other possible worlds. So, Better Semantic Markers will not help us in

31 'Being monarch-looking' refers to the property of looking the way monarchs look *in the actual world*. If the property referred to the different ways monarchs look in different worlds, there would be no world at which monarchs do not instantiate the property 'being monarch-looking'.

32 Notice that this solution is inspired by Fodor's asymmetric dependence theory; content depends on the causal relation holding in worlds where some of the current causes fail to exist.

determining content for the same reason the Crude Causal Theory did not work out: the fact that misrepresentation is possible shows that content cannot be determined by what causes a certain mental state. And this claims holds here and in other possible worlds.

The reason Prinz theory faces a misrepresentation problem at that particular point and not earlier, is that if we focus on the actual world, he has an adequate reply: only the first cause (the incipient cause) determines content. So (if we assume that concepts are always firstly caused by instances of their referents), he has a way of distinguishing misrepresentations from true representations. In contrast, when he appeals to causal relations holding in other possible worlds in order to determine the content of the concept in the actual world, his theory yields the wrong results. In other possible worlds, anything can cause my mental state.[33] Consequently, even if Monarch means *monarch* and not *butterfly*, if we move to worlds where butterflies are not monarchs, we will probably find that some butterflies (which are not monarchs) still cause tokens of the concept and in some of this worlds they do not. As a result, the fact that X and not Y causes tokens of a concept C at those worlds where X and Y are not coinstantiated cannot help to determine content.[34]

Therefore, pace Prinz, I think Prinz' theory cannot solve the indeterminacy problem.

ambiguity Secondly, not only the nomological condition, but also the 'incipient cause' condition runs into problems.

There are at least two kinds of counterexamples that Prinz has not appropriately addressed. First of all, according to Incipient Theory, (non-deferential) concepts can never have ambiguous contents. Suppose I have a concept C that I equally apply to beeches and elms, and suppose I have never heard about these trees, nor do I intend to defer the fixation of meaning to experts (so, suppose this concept C is not deferential). In that case, my concept would either mean beech or elm, depending on

33 Furthermore, in that case, he cannot modify Better Semantic Markers so that only the first cause in other possible worlds is relevant. That would surely be too strong a condition; even if we grant that in the actual world my concept TREE was first caused by a tree, there are many possible worlds were trees are not the incipient cause of my concept TREE.

34 In a previous version of the theory, he offered a slightly different condition. In Prinz (2000, p. 13) he claims that X nomologically covaries with Y iff (1) Xs cause Ys in all proximate nomologically possible worlds, and (2) *when they do so, they do so in virtue of being Xs.* Accordingly, one might think this previous version avoided the problem of indeterminacy I am pointing out, because when Prinz appeals to what happens in other possible worlds, he has a way of distinguishing the right causes from the wrong causes by appealing to the relation *in virtue of*.
But, of course, an obvious reply is that this relation of *in virtue of* is doing all the work and should be specified further. If this relation is not explained, one might worry it is presupposing precisely what it is trying to explain, namely that Y means X (see below the section on 'circularity'). Think about it in the following way: if one were allowed to appeal to X causing Y in virtue of being an X, then nothing like incipient causes or nomological covariance would be be required. One could just say that Y means X iff Xs cause tokens of Y in virtue of being X. That would surely be a vacuous naturalistic theory. In this approach, the notion 'in virtue of' seems to merely label the relation we are trying to explain, rather than offering an explanation.
Indeed, Prinz (2000, p. 13) admits that this notion should be explained, and claims that Xs cause Ys in virtue of being Xs when (1) when an a that is X causes Y, if a were not X, it would not cause Y or (2) when an a that is X causes Y, there is no other nomologically sufficient cause of Y. However, this way of cashing out the relation *in virtue of* shows that this previous version of the theory also suffers from the indeterminacy problem. Even if my concept MONARCH means *monarch,* some butterflies cause tokens of this concept in the actual world and in other possible worlds.

the entity that first caused it. That is an implausible result, for various reasons. On the one hand, if I have always consistently and repeatedly applied a concept C to two entities, the intuitive result is that this concept is ambiguous. Secondly, whether an elm or a beech was the first cause seems to be a matter of luck, but the content of my concept C does not seem to depend on such a chancy event. Thirdly, it clearly seems that, as a matter of fact, some of our concepts are ambiguous, and language does not seem to be required for having them (Millikan, 2000). In this case the strictness of INCIPIENT CAUSE makes it difficult to account for these cases where meaning is disjunctive.[35]

CIRCULARITY Finally, I would like to raise a general worry concerning this sort approach. A striking problem with BETTER INCIPIENT THEORY is that (as we saw when discussing Fodor's Asymmetric dependence theory in 1.2.4) we lack a (non-intentional) justification of why 2 should hold. Of course, it is true of many of our concepts that in the most proximate worlds their referent still causes them, but this is usually *explained* by appealing to the fact that concepts mean what they mean. In other words: Why do monarchs in most proximate worlds cause my concept MONARCH? Well, a plausible and intuitive explanation is that this is true precisely because MONARCH means *monarch*. That means that the intuition that 2 is on the right track comes from the fact that MONARCH means *monarch*; so in premise 2 we are assuming what we are trying to explain.

Let me put the point in a different way. The truth of counterfactual statements is usually thought to be grounded in (categorical) properties and relations holding in the actual world (at least, that seems to be a usual assumption of naturalist accounts). For instance, consider the following counterfactual: *If Obama had not won the elections, Romney would have become the U.S. president.* Unless we are modal realists (Lewis, 1986), we will probably think that this counterfactual is true because of certain properties and causal relations holding in the *actual* world. Now, the general problem with attempts to naturalize content by appealing to counterfactual conditions such as Prinz's is that there is always the worry that the truth of the counterfactuals they are appealing to might be grounded on the intentional relations holding in he actual world that they are trying to explain. So, unless they specify which properties and relations in the actual world account for the truth of these counterfactuals, the naturalistic credentials of this account will be dubious. In order to provide a full characterization of a concept and its content, one should establish in virtue of which non-intentional property this nomological relation holds. The fact that no such characterization is provided, suggests that these accounts may rely on the intuitions that they are trying to explain.

Therefore, Prinz's account faces a wide range of daunting difficulties that suggest that BETTER INCIPIENT THEORY (or INCIPIENT THEORY) is probably not the right naturalistic account of conceptual content. Let us move to teleosemantic theories of conceptual content.

35 I focus on non-deferential concepts because Prinz (2002, p.254-5) has provided an interesting solution to this problem for deferential concepts: since in the case of deferential concepts there is a community involved, there might be different incipient causes for the same concept (Prinz, 2002, p.254-5).
Even if we granted that this proposal can work for deferential concepts, this solution is surely not available to the case of non-deferential concepts.

5.2.3 *A Top-Down Teleosemantic Account of Conceptual content*

In this section, I will consider several ways in which teleosemantic ideas have been brought to bear on the question of conceptual content. My primary goal is to show that the employment of teleosemantics within the conceptual domain is much more complex than it has been usually thought. I will show that there are different strategies a teleosemanticist can take, and each option carries with it several difficulties.

Let us start by discussing Papineau's attempt to develop a teleosemantic account of concepts. Afterwards we will focus on Millikan's own approach.

5.2.3.1 *Papineau (1998)*

Papineau's own naturalistic account of representation has mainly focused on thoughts and desires, rather than the representational capacities of cognitively unsophisticated organisms. He thinks it is possible to provide a naturalistically acceptable recipe for attributing certain thoughts to organisms, and that this perspective will enable us to specify whether other organisms have representational states as well. Thus, he adopts what he calls a 'top-down strategy', in contrast to the standard 'bottom-up strategy'. Papineau's main reason for adopting the former is that simple organisms are poor starting points for a naturalistic theory ('Frogs are bad examples'). We lack clear pre-theoretical intuitions concerning their representational states, and for this reason it might be very hard to assess which theory yields the right results.[36]

Papineau's most recent and elaborate proposal is intended as a combination of some of Millikan's and Neander's insights. I think that his theory can be best introduced with an example. Suppose someone has the desire to eat chocolate. Then,

> If we now assume that the biological purpose of belief S is to be present in those circumstances where the behavior it prompts will satisfy the desires it is combining with, then it follows that the content of the belief S is [that there is chocolate]. (Papineau, 1998, p. 10)

More generally, if an organism has a desire with a certain determinate content, then there is a belief that has the function of combining with this desire and prompt an action that satisfies this desire. Now, the content of this belief is determined by the circumstances that are required for the desires to be fulfilled. This is how the contents of beliefs are determined in organisms with a belief-desire psychology. As in Millikanian teleosemantics, descriptive content is determined by the conditions that are required for the system that consumes the representation to act successfully (i.e. to satisfy a desire).

However, notice that on this account the content of beliefs is determined by appealing to the content of desires. So, for the theory to be fully naturalistic, a reduction of the content of desires must be provided. At this point is where Papineau appeals to Neander's proposal. The idea is that the content of the desire-state is the specific effect the desire is supposed to bring about:

36 Indeed, Papineau (1998) provides a substantive explanation of the diversity of intuitions at that point: he argues that intuitions concerning the representational content of frogs and other unsophisticated organisms are blurry because there is no fact of the matter. Since frogs lack a belief-desire psychology, their states probably lack a determinate content. If that were true, it would lend further support to a top-down strategy.

> "My suggestion is that the teleological theory should identify the satisfaction conditions of a desire as the result which is the desire's specific function" (Papineau, 1998, p.14).

But, do desires have specific effects? And, how can we ascertain the specific effect of a given desire? Papineau suggests that we should divide the problem into two.

First, any given desire has many proximal effects (sending electrical impulses to the muscle, moving your arm,...). Papineau claims that from all the effects a desire is supposed to produce, in order to find its *specific* function we need to go far enough along the chain to reach results which do not depend on which beliefs the desire happens to be interacting with (Papineau, 1998, p. 15) -in the same way Neander claimed that pumping blood is the specific function of hearts because apparently this result do not depend on any other trait (see 3.3.1). There is, however, a different source of indeterminacy. Even if we exclude moving one's arm and sending electrical impulses, there are many effects that one performs by eating chocolate: food is digested, health is preserved,... Papineau claims that here is where we need to employ Neander's appeal to proximal causes. The specific function is the most proximal effect in this chain, which according to Papineau is something like *chocolate is ingested.*

In a nutshell, Papineau's solution consists in deriving belief's content from the desire's content, and then provide an account of the content of desires by considering the desire's effect that (1) does not depend on the belief the desire is actually acting in concert with, and (2) is the most proximal effect that satisfies condition 1.

Let us now assess whether such an account can be satisfactorily worked out and whether it can solve the problems of previous approaches.

PROBLEMS WITH PAPINEAU'S (1998) THEORY Probably every step in Papineau's proposal faces serious objections. On the one hand, there are certain worries with the functional talk in Papineau's proposal. Since he adopts an etiological account of function and thinks that beliefs and desires have them, he seems to be assuming that every contentful belief and desire have been selected for, what is extremely implausible (Devitt, 1991). Furthermore, I argued in chapter 3 that representational content cannot depend on the function of the representation itself (see 3.2.6). This is an issue I already discussed extensively, and since there are more pressing problems with Papineau's proposal, I suggest to leave this question aside.

Now, consider the idea that the content of beliefs is determined by reference to certain desires. A first intuitive difficulty with this suggestion is that it seems that I believe many things concerning certain entities which I do no have any desires about. I might never have had any desire concerning stars, penguins or oaks but I have had many thoughts about them. That looks like a very common phenomenon. How could this proposal deal with beliefs (and concepts) that are not accompanied by any desire? There are two options available, but I doubt any of them can utterly succeed.

On the one hand, one could claim that the concept STAR is composed, say, by BRIGHT LIGHT and SKY, and accordingly the content of STAR is determined by the content of the composing concepts. But, even if we were to grant that there are some basic concepts whose composition

can provide the content for the rest of our conceptual apparatus, this view would be committed to hold that we have had desires concerning the referents of all these basic concepts, what seems very implausible (why should we believe basic concepts are less problematic concerning our desires towards them than less basic ones?). Indeed, it is extremely plausible that many of the basic concepts refer to entities we have never had desires about (e.g. PURPLE, FLOOR, OAK). The original worry just reappears at that point.

On the other hand, a second possible strategy for dealing with the content of concepts that refer to entities I have never had any desire about is to hold that the content of beliefs is determined by *possible* desires. For instance, one could claim hat *if* I were to have desires concerning stars, my STAR-involving beliefs would lead to successful actions only if there were stars. According to that proposal, the content of my belief (or the content of the concepts that constitute my belief) is determined by the state of affairs that would lead to the satisfaction of the desires I could have. This suggestion, however, also faces striking difficulties. A first problem is that if we do not restrict the set of possible desires that we should consider when assessing the content of a given belief we might get very counterintutive results. As Neander (2012) points out:

> Consider also the following scenario. The desire to detect phlogiston might tend to cause oxygen detection (i.e., oxygen which is mistaken for phlogiston). Further, being a successful scientist might contribute to one's fitness, and seeming to have detected phlogiston by really having detected oxygen might contribute to being a successful scientist. Thus it would seem that, in this scenario, and according to Papineau's theory, PHLOGISTON means oxygen.
>
> Furthermore, notice that since we are told to consider all desires that could be satisfied with this belief, we would get to the conclusion that PHLOGISTON means phlogiston and oxygen and any other thing that could satisfy my desire for fame. So, I take it that we have to find out a way of restricting the set of desires that are relevant. The problem, of course, is that I do not see any principled way of picking out this privileged set of desires without appealing to the content of beliefs.

So, the fact that many of my beliefs have never been combined with a desire raises a serious problem for the theory. Insisting on the idea that the basic constituents of our beliefs have always been combined with certain desires or appealing to possible desires does not seem to provide plausible solutions to these worries.

Indeed, there is a related problem that concerns all beliefs, even those that have been combined with certain desires. Think about those beliefs that in fact have lead us to behave in a certain way because they have interacted with certain desires. Papineau claims that the content of these beliefs depends on the content of the desire they have been combined with. But, in general, beliefs and desires can be conjoined in a vast number of different ways. The relation between beliefs and desires is not one-to-one but many-to-many; as a consequence, unless there is a (non-intentional) way of restricting the set of desires that can be combined with a given belief, the theory will have extremely

counterintutive results. In other words: nothing ensures that the belief CHOCOLATE IS TASTY has been combined with a desire whose satisfaction conditions were the presence of tasty chocolate (rather than, say, expensive chocolate).[37]

Let us now focus on the content of desires. First of all, Papineau claims that the desires have been selected because they give rise to their satisfaction conditions. Now, this process of selection is either filogenetic or ontogenetic. Either way, the account runs into serious troubles (Devitt, 1991). In the first case, Papineau's theory cannot account for irrational desires, like the desire for saturated fat or drugs. Surely, these desires did not increase the chances of reproducing. On the other hand, if the process of selection is supposed to take place during the lifetime of the individual (and 'fitness' corresponds to something like 'being psychologically rewarding'), then it is dubious that most desires have been psychologically rewarding in that sense. Many of our desires remain unsatisfied.

But there are also problems with the desire's content. According to Papineau, even if my desire causes me to move my arm in such and such direction so as to reach the chocolate and that effect explains the success of my action, such movement cannot qualify as the content of my desire because it depends on the presence of certain beliefs. According to him, the content of my desire is determined by an effect that does not depend on any particular belief I have, what he calls the desire's 'specific effect'. However, it is dubious that there is any specific effect of a desire that it can bring about without the aid of a suitable belief. For instance, my desire for chocolate causes me to ingest chocolate only because I have certain beliefs: the belief that there is chocolate in the world, that chocolate is not going to kill me, that chocolate is tasty,... More generally, the notion of 'specific function of a desire' does not seem to pick out any property of desires (compare with the problems of Neander's notion 'specific function', discussed in 3.3.1). This issue is crucial because, unless Papineau provides a naturalistically acceptable account of the content of desires, his account will interdefine the content of beliefs and the content of desires. As a consequence, the naturalistic credentials of his proposal would be jeopardized.

Finally, and more importantly, the virtues and main intuitions in favor of Papineau's account (the fact that the content of mental states depends on the conditions that has enabled successful actions in the past) has already been captured and developed in the teleosemantic account I have put forward in the first part of this dissertation. And, obviously, I think that the teleosemantic account I have defended has many advantages over Papineau's view. On the one hand, it does not have any of the problems of Papineau's account. On the other, it provides a unified account of the phenomenon of representation in cognitively unsophisticated animals, animal signaling, sub-personal states and human conscious states, which is very unlikely to follow from Papineau's top-down approach to the phenomenon of representation. Consequently, I think that the account I suggested keeps the advantages and avoids the main difficulties of Papineau's approach. A full defense

37 At a certain place, Papineau (1993, p. 62-3) appeals to a distinction between 'normal' and 'special' functions of beliefs and desires. But, of course, this is just a way of labeling the problem; unless a naturalistic account of this notion of normality is provided, nothing has been gained. And notice that the notion of 'Normal' defined in 2.2.2.1 is of no use at that point, because Normality was defined in terms of natural selection, and probably concepts and beliefs are not selected for in that strong sense.

of how Third Teleosemantics applies to the conceptual domain will be offered in the next chapter.The upshot, I think, is that Papineau's top-down strategy is unlikely to provide the right theory of content.

Despite this negative result, Papineau's original proposal illustrates the fact that there are many ways of developing a teleosemantic account of the content of cognitive representations. Indeed, Papineau intended his account to resemble very much Millikan's view, on which I have based my own proposal. Unfortunately, I have just argued that taking a top-down strategy is probably not the best option here. Let us now move to Millikan's approach.

5.2.4 *Millikan on Concepts*

In some of her work, Millikan has extensively argued for a particular view on the nature, structure and content of concepts. Given the impact of her work on the literature on concepts and the fact that I have been following her on central questions concerning the phenomenon of representation, I would like to consider in some detail her view on that matter.

5.2.4.1 *Nature*

First of all, it is worth pointing out that Millikan does not aim at defining all concepts, but only a subset of them, what she calls 'substance concepts'. As a first approximation, substance concepts are concepts that refer (or are supposed to refer) to substances. More precisely:

Millikan Concept Substance concepts are abilities to reidentify substances (Millikan, 1984, p. 318; 2000, p. 51).

In order to properly understand this claim, let us define each of the notions involved in this definition.

abilities As Millikan (2000) makes clear, her answer to the question about the nature of concepts is that they are abilities and she provides a detailed characterization of what should be understood under 'ability'.

Nevertheless, as I pointed out earlier (see 5.1.1.3), Millikan intends her theory of concepts as abilities to be compatible with the claim that they are mental representations (cfr. Lawrence and Margolis, 2011; Millikan,2013).[38] So, even if conceiving concepts as abilities is not the same as regarding them as mental representations, Millikan thinks the two views are compatible.

substances Millikan takes the notion of 'substance' from a broad Aristotelian tradition, but gives it a specific meaning. Her definition goes as follows:

> Substances are those things over which you can learn from one encounter something of what to expect on other encounters, where this is no accident but a result of a real

38 Here are two quotations where that is clear: "There is another tradition that treats a theory of content as part of a theory of cognition by taking a concept to be a mental word. If one takes it that what makes a mental feature, or a brain feature, into a mental word is its function, then this usage of "concept" is not incompatible with my usage here (Millikan, 2000, p. 2). "(...) we also can think of substance concepts as corresponding to mental representations of substances, say, to mental words for substances *but qua meaningful* (Millikan, 2000, p.13).

connection. (Millikan, 2000, p.15; see also Millikan, 1984, p.275)

There are two different aspects of this definition, one metaphysical and the other epistemological. Let me explain them very briefly.

On the metaphysical side, there are two conditions that must hold in order for a group of entities to form a substance. First of all, a substance is typically formed by a number of entities that have many properties in common (so that you can learn many things about the properties of any member by focusing on the properties of other members). Cats, for instance, tend to share many properties: they are four-legged, they have fur, they have a mustache, they miaow... Secondly, the fact that the entities that form a substance share all these properties cannot be mere luck, but must be the result of some kind of real connection (Millikan, 2005, ch. 6). For instance, the fact that most cats have a mustache is no coincidence, since they all have a common ancestor with a mustache and there are causal processes that account for the permanence of this trait.

Now, the epistemological aspect follows from this metaphysical description: since different instances of the same substance tend to share the same properties due to a real connection between them, inductions and inferences over substances and their properties are very likely to be true. If a cat has four legs (and assuming that this is one of the projectable properties), probably any other cat will also have four legs.

Consequently, since the epistemological aspect is entailed by the two metaphysical conditions, I suggest to define 'substance' by merely focusing on the metaphysical side.[39] Hence:

SUBSTANCE A set S of entities forms a substance if an only if

1. Members of S share many properties in common
2. Condition 1 is satisfied in virtue of some underlying causal process.[40]

Notice that (First-Order) Reproductively Established Families (REF) count as substances, but there are some substances that do not count as REF. Remember the definition suggested above:

FIRST-ORDER REF A set of individuals $d_{1,}, d_{2,}, d_3, ..., d_n$ form a reproductively established family D iff

1. There is a set of properties $F_{1,}, F_{2,}, F_3$ such that $d_{1,}, d_{2,}, d_3, ..., d_n$ tend to instantiate a high number of these properties
2. For any d, the fact that d's ancestors had $F_{1,}, F_{2,}, F_3, ..$ in part causally explains why d has $F_{1,}, F_{2,}, F_3, ..$

Condition 2 in FIRST-ORDER REF is one of the causal processes than can satisfy condition 2 in SUBSTANCE, but there might well be other causal processes that fulfill the latter condition. That shows that all First-order REFs are substances, but there are more substances than FIRST-ORDER REF. The chief difference is that FIRST-ORDER REF requires

39 Of course, this is not intended to imply that the epistemological side is irrelevant. Indeed, this definition of substance is interesting at all because it groups together a set of entities in a way that is relevant in the explanation of certain cognitive abilities (Millikan, 2000, p. 26). My claim is rather that, whatever the motivations for these notions are, one can provide an adequate definition by merely relying on the metaphysical side.

40 There is a close connection between this notion of substance and Boyd's notion of 'Homeostatic Property Cluster', exploited in Martinez (2010).

that entities that belong to a REF be causally connected to each other, while SUBSTANCE does not include this condition. Gold or Water, for instance, are not REF, but what I called 'Instance-types' (see 3.2.4), because different pieces of gold (or water) are not related to each other by some causal process (see discussion of Martinez's view in 3.1.2). Nonetheless, they qualify as substances, because all pieces of gold share many properties (solidity, reflectance, color, melting point,...) due to an underlying causal process (which apparently has to do with its essential properties and probably the laws of nature).

This definition of 'substance' is accompanied with a *sui generis* classification of kinds of substances. Millikan mentions three different kinds of substances: on the one hand, she distinguishes natural kinds (human beings, stars, electrons), stuffs (water, milk) and individuals (Barack Obama, Walt-mart). In turn, each of these groups can be divided into those substances that are historical (human beings, peanut butter) and those which are ahistorical (star, water).[41] Since concepts are abilities to reidentify substances, any of the entities we can think of (any entity we have a concept of) has to fall within one of these categories.

But, one might worry, it is not clear that our cognitive capacities are restricted in that way. If concepts are abilities to reidentify substances, then can we only have concepts of substances? What about concepts such as WHITE, SQUARE or TALL? Two things should be said in that respect.

First, in some places she seems to be suggesting that properties can be considered as substances as well:[42]

> Squares and cubes of material are things that one can learn to recognize and about which one can learn a number of stable things, such as how they fit together, how they balance, that their sides, angles, and diagonals are equal and so forth. As Cangelosi and Parisi remark (correcting me), white gets dirty easily and, I now add, shows up easily in dim light, stays cool in sunlight but also tends to blind us, and so on (Millikan, 1998, p. 92; see also Millikan, 2000, p. 27).

Properties such as *being white* and *being square* fulfill 1 and 2 in SUBSTANCE, so on the Millikanian way of understanding substances there is no reason for thinking that they cannot qualify as such.[43]

Secondly, Millikan's goal is to provide a theory of substance concepts (Millikan, 1998, p.56). So she is aware that we might possess some concepts that are not substance concepts in her sense. Thus, if we found a set of concepts referring to some entity that did not fulfill 1 and 2 in SUBSTANCE, Millikan would probably reply that these are not the kind of concepts she is talking about.

41 Millikan does not mention the category of *ahistorical individuals*, but this group seems to be suggested by her classification (and, in fact, she talks of 'historical individuals', what seems to indicate that there are ahistorical ones (Millikan, 2000, p.24). Some examples might be the Sun or God.

42 In contrast, at other places she seems to think that they are not: 'I propose that individuals, basic-level kinds and stuffs have something in common that makes them all knowable in a similar way, and prior to properties' (Millikan, 1984, p.56).

43 Let me add that in Millikan (1984, ch. 15-17) she provides a different definition of substances and properties. On this definition (which may or may not be compatible with her official view since then), substances and properties are interdefined. What counts as a substance at a given time depends on what counts as a property and vice versa. A detailed discussion of these issues would lead us too far away from our present concerns.

REIDENTIFICATION The notion of reidentification is, I think, the less clear from the three that figure in the definition. As a first approximation, a subject reidentifies a substance when it is able to identify the same substance in different situations. A passage where this idea seems clear:

> Recognizing Mama by smell certainly is not classifying her nor is it conceiving of her as whatever bears that smell. It is more accurate to imagine it as a tokening of the mental term "Mama" in response to a smell. (Millikan, 2000, p.81)

Reidentifying involves tokening the same representation when the same substance is present, or (more akin with Millikan's own terms) the ability to identify the same substance in different occasions.

A different way of expressing the same idea is that having a concept requires being able to track a substance in different environments and through different media. For instance, I can track my cat by smell, by touch or by sight, in different positions and environments. I have an ability to reidentify my cat if I am able to track the cat in different circumstances.

However, Millikan emphasizes that merely tokening the same mental word when one is confronted with the same substance S does not suffice for reidentifying S (Millikan, 2000). Reidentifying a substance requires more than just tracking it in different occasions. I also need to recognize that the substance I am tracking is the same one. That is, I reidentify my cat in different occasions only if (1) I track it in different circumstances and (2) I know it is the same substance that I am tracking. Unless I somehow ascertain that it is a cat again, I will not be able to use the same kind of information that I learned in previous situations. I must be aware of the fact that the substance I met previously is the same one I am presently tracking.

Now, what it is to know that one is tracking the same substance in different occasions (that is, condition 2 above) is something that Millikan has spent a lot of time explaining. In brief, the idea is the following: if one uses a mental term M_1 when tracking a substance at t_1 and a mental term M_2 when tracking the same substance at t_2, one knows that $M_1$and M_2 refer to the same substance iff one is disposed to use any information associated with one of these terms as if it were also information associated with the other term. In other words, if I am disposed to use any information I have about Tully when I am confronted with Cicero and vice versa, that means that I know that TULLY and CICERO refer to the same entity.[44]

Hence, the process of reidentification can be defined in terms of knowingly tracking a substance at different times:

REIDENTIFICATION A subject reidentifies a substance S at different times $t_1 ... t_n$ iff:

1. The subjects tracks S at $t_1 ... t_n$

44 The same idea can be expressed in terms of functions and proper functioning of devices: "An act of correct identification [i.e. knowing that two terms refer to the same substance] is performed by an interpreting device that uses these icons jointly in order to perform a proper function where the Normal explanation for proper performance of this function makes reference to the fact that the real value [i.e. the referent] of these two elements is the same. That is, the interpreting device will be able to accomplish what good it does Normally only *because* these elements map the same. The act of identifying operates upon pairs of intentional icons. But in so doing it identifies variants in the world."(Millikan, 1984, p. 242)

2. At any time $t_1 ... t_n$, the subject is disposed to use the information about S gathered previously.

Reidentifying basically consists in tracking the same substance at different moments and being disposed to carry over the information gathered at one time to the future encounters with the substance.

We have already explained condition 2; what about 1? What it is to track a substance? Millikan (1998) describes three central ways of tracking a substance. First, there is perceptual tracking, by means of which a subject is in perceptual contact with a substance and she can assess that she is confronted with the very same substance. Certain inborn mechanisms like the ones underlying perceptual constancies may allow subjects to perform this task (Millikan, 1984, p. 255). Secondly, there is conceptual tracking, which makes use of higher-order cognitive abilities and enables subjects to track substances in different places and occasions. Most substance concepts require both perceptual and conceptual tracking. Finally, Millikan controversially claims that we can track substances by means of language. We can label this third kind of ability 'linguistic tracking'. All these claims will be discussed below.

Now we are in a position to understand Millikan proposal on concepts: she thinks that concepts are abilities to reidentify substances, that is, they are abilities to track certain kind of entities at different occasions, such that the information gathered at one encounter can be reliably carried over to new encounters. In a nutshell, a subject has a concept of substance S when it is able to identify S in different occasions and she is a aware of the fact that she is identifying the same substance.

5.2.4.2 *Structure and Content*

In the previous section I have described Millikan's view on the nature of concepts. Let us consider now her approach to the structure and content of concepts.

CONCEPTIONS Remember that when we were discussing the structure of concepts, we saw that there is a wide range of empirical data suggesting that concepts are usually accompanied with prototypes, exemplars and theories. This conceptual network enables us to reliably identify the entities our concepts refer to. What is the role of prototypes, exemplars and theories in Millikan's account? The information attached to a concept (prototypes, exemplars and so on) is what she calls 'conception'. The distinction between concepts and conceptions is a key topic in her writings.

Whereas concepts are abilities to reidentify substances, conceptions are constituted by the ways a subject has of reidentifying a substance. Conceptions are constituted by the set of knowledge one has about a certain entity:

> The 'conception' one has of a substance, then, will be the ways one has of identifying that substance plus the disposition to project certain kinds of invariances rather than others over one's experiences with it (Millikan, 1998, p. 90; 2000, p. 12).

Within the conception a subject has of a substance, Millikan distinguishes the knowledge one has of the *kind* of information that can be

gathered (e.g. color, size, melting point,...) from the information itself (e. g. blue, round, 100º). She calls the former 'template'. Since these notions are going to pay an important role in the next chapter, let me explain this distinction in some detail.

Generally speaking, any entity belongs to different substances. For instance, there are many substances my cat Fluffy is a member of: he is of course a cat, but also an animal, a housecat and *Fluffy* (an individual). So, if I track Fluffy at different moments, how do we know whether I am developing a concept of the natural kind *cat*, or the kind *animal* or the individual *Fluffy*? Millikan claims that the answer depends on the kind of invariances I am disposed to project to other encounters. For instance, if in my encounters with Fluffy I merely retain the fact that it has four legs, a mustache, it is flurry and so on, I will probably be developing a concept of cat. If, instead, I try to retain its name, the particular color of its skin, its particular smell, where it lives, and so on, then I am developing a concept of an individual (Fluffy) rather than a concept of a natural kind. This is what Millikan calls a 'template'; a template is a rough idea of what kinds of properties can be learned from a substance. For instance, in order to develop a concept of an animal such as elephant I need to use a template composed of *color, number of legs, size,...* The template is the kind of questions that it is adequate to ask. It makes sense to ask how many legs elephants or chickens have, but not how many legs gold or silver has. Similarly, gold and silver have a determinate melting point, but elephants and chickens have not. So the template used when developing a concept of gold or silver differs from the template used when developing a concept of elephant or chicken.

Obviously, the same template can be used in producing concepts of different entities (i.e. the same questions can be reasonably asked about different substances), so we do not need to learn a different template for every substance. For instance, it is likely that the substance concepts of many mammals have the same template, since we classify mammals by means of asking the same kind of questions. The information we attach to DOG is different from the information we attach to ELEPHANT, but the template (the questions we ask) may still be the same. In contrast, in order to classify different kinds of clouds, political parties and colors we surely use different templates as well as different information. Conceptions include the templates and the specific information that fill them.[45]

Now, Millikan's distinction between concepts and conceptions is central to her account, because her particular view on the relation between them opposes most traditional philosophical views on the matter and a big part of cognitive science. In particular, there are two controversial theses that Millikan endorses:

1. Conceptions do not determine the extension of concepts.[46]

45 Millikan also includes some forms of knowing-how (i.e. procedural knowledge) within the conception (Millikan, 2005, p. 69). Any means a subject employs in order to identify a substance is part of the conception.

46 For instance, after reviewing some papers that misinterpreted her view, she added: 'None of these claims is what I had in mind when rejecting descriptionism. The descriptionist holds that the conception one has of a substance determines its extension'. (Millikan, 1998, p. 91). She also talks of 'conceptionism' that is, "the view that the extension of a concept or term is determined by some aspect of the thinker's conception of its extension, that is, by some method that the thinker has of identifying it (Millikan, 2000, p. 42).

2. There is no necessary connection between possessing a concept and possessing a particular conception.

Notice that, using the terminology set up earlier, these are just nominal variations of Semantic Atomism and Conceptual Atomism:

Semantic Atomism The referential content of a concept is not determined by its relation to other concepts.

Conceptual Atomism For any standard concept C, there is no particular set of non-logical concepts S, such that a subject needs to posses S in order to possess C.

So Millikan's theory is a paradigmatic example of atomism with respect to content determination and atomism with respect to the structure of concepts.

After describing Millikan's views on the nature and structure of concepts, let us move to the discussion.

5.2.4.3 *Are Concepts Abilities to Reidentify Substances?*

What are Millikan's main arguments in favor of her view on concepts? In this section I will discuss some of the specific arguments she gives for her proposal, as well as more general considerations about the theory.

substances First of all, concerning the very existence of substances, Millikan (1998, p. 56) brings forward some empirical arguments based on developmental psychology. Her purpose is to argue that there is a sense in which natural kinds, stuffs and individuals have a common structure (defined in Substance). For instance, she claims that names of individuals, names of basic-level kinds and names for stuffs ('milk', 'juice') are learned first between one-and-a-half and two years of age. Only later infants develop names for abstract objects and adjectives (Gentner, 1982; Markman, 1991). If we assume that conceptual development in children roughly depends on objective similarities between entities, these findings suggest that there is a common structure between kinds and stuffs. Similarly, Carlson (1998, p.68) argues that, from the grammatical point of view, mass terms (i.e. stuffs like 'gold', 'water') and kind terms ('lion', 'pencil') have distributional and semantic properties in common with proper names, so they work much in the same way in language.[47] These empirical arguments are also supplemented with some metaphysical considerations. Since I will not focus on the metaphysics of substances and, in any case, that issue will not affect any of the points I would like to discuss, I think we can leave the metaphysical question aside. Nevertheless, let me add that it is not unreasonable to think that entities classified as substances by Millikan form real kinds. *Prima facie*, entities that satisfy Substance seem to be good candidates for constituting real joints in nature.

With respect to the relation between concepts and substances, the most important argument for the view that concepts are abilities to track *substances* appeals to evolutionary considerations. She argues that concepts have been useful to organisms in evolution because they have allowed organisms to gather information about certain entities that could be employed in future situations. Possessing a mechanism

47 Let me stress that Millikan does not seem to put much weight on these arguments and, at some point, she even seems to be dismissing them as irrelevant (Millikan, 1998, p. 94).

that produces representations of substances seems to be clearly advantageous in evolution because it allows subjects to gather information that can be reliably projected. This is one of Millikan's main arguments in favor of the view that concepts are primarily abilities to track substances. In fact, I will suggest that this argument should be regarded as the most important reason in favor of this understanding of concepts.

I think that Millikan's treatment of *substances* is mainly right, but my own approach takes this view a bit farther away. When we were discussing the indeterminacy problem in 2.3.3, I already argued that substances (either REFs or Instance-Types) are the most plausible candidates for figuring in the content of many mental states. The reason was that, very often, it is the presence of a substance what provides the least detailed and most comprehensive Normal explanation of the success of the consumer system. Let me recover the examples I discussed: even in the case of frogs, it is the presence of a *fly*, rather than the presence of a *small, nutritious insect,...* what provides the least detailed and most comprehensive Normal explanation of the success of the consumer system. So the content of the frog's mental state is *there is a fly*. Similarly, I argued in 3.1.2 that it is the presence of *water*, rather than the presence of a *tasty, refreshing and transparent thing* what provides the least detailed and most comprehensive Normal explanation of the success of the consumer system. And so on. Therefore, in my view it is a general fact about representational systems that they primarily refer to substances. Concepts are just a particular case of this general truth.

In a nutshell, what I am trying to claim is that in the first part of the dissertation I already argued that most representational systems (even the representational systems of cognitively unsophisticated organisms) represent substances. In that respect, concepts fit this general schema. Consequently, I completely agree with Millikan that concepts primarily represent substances, but I add that there is nothing specific about concepts here. Most representations (specially those of simple organisms) are about substances.

CONCEPTS AS ABILITIES Remember that Millikan extensively argues that concepts are abilities, but at the same time holds that this claim is in agreement with the thought that concepts are mental representations. Now, I think that there is a clear tension between these two theses; it is not obvious that the idea that concepts are abilities is compatible with the idea that they are mental representations (Lawrence and Margolis, 2011). For one thing, I pointed out earlier that concepts compose (we can form complex concepts out of two or more concepts, as in BLACK SWAN), but it is not clear what it would be for an ability to compose (Fodor, 2008, p. 45) As Gauker (1998, p. 71) claims:

> One would not say that concepts are action schemata or that concepts are structures composed of theories, because such things cannot go together to form thoughts in the way words go together to form sentences. For the same reason, one would not say, with Millikan, that concepts are abilities to reidentify.

We mentioned earlier that another difficulty of this approach is that it seems ill-suited for explaining the role that concepts play in mental processing (inferences,..). Furthermore, while it is a platitude that concepts are meaningful, it is not straightforward in what sense abilities

can have meanings. Do abilities represent the world as being a certain way? It is unclear how they could.

As I pointed out earlier, Millikan thinks that her view on the nature of concepts is compatible with a view according to which concepts are mental representations. However, all these different properties of abilities and mental states suggest that the two approaches are not obviously in agreement. Be as it may, I have already provided some reasons for favoring the view that concepts are mental representations (see 5.1.1.4), so I will keep talking of concepts as mental states.

SUFFICIENCY A general difficulty with MILLIKAN CONCEPT is that it does not seem to specify *sufficient* conditions (or even *nearly sufficient* conditions) for a mental state to qualify as a concept. There are several mental states/abilities that satisfy MILLIKAN CONCEPT but should not be classified as *conceptual representations*. For instance, suppose a rat has been conditioned to press a bar every time a certain light is on. Thus, whenever it perceives a light being on, it knowingly identifies it as the same light that was present at earlier moments and acts accordingly. It is usually assumed that, in operant conditioning, this whole process of recognition and action can take place without concepts. If MILLIKAN CONCEPT was taken as specifying sufficient conditions, it would entail that organisms that are able of operant conditioning are endowed with concepts. But many organisms that are apt to some forms of conditioning, like toads or salamanders, surely lack concepts (Allen, 1998, p.66). The worry is even more pressing when we consider the fact that most representational systems in fact represent substances (see above). Neither the act of *reidentification*, nor the fact that an organism is reidentifying *substances*, distinguish conceptual representations from other kinds of states. So MILLIKAN CONCEPT only offers a partial characterization of concepts.

At root, this issue points at a broader question: Millikan's theory does not address two central aspects of conceptual representations, namely compositionality (Martinez, forthcoming) and its essential connection to thoughts. The fact that concepts compose and form thoughts is one of their defining features. This is not so much a mistake, but a particular aspect of a teleosemantic theory of concepts that needs to be addressed. I will elaborate on that issue in the next chapter.

PROPOSITIONAL/SUBPROPOSITIONAL CONTENT A related point is that Millikan's sender-receiver model, as well as her work about representations in cognitively unsophisticated organisms, is based on the assumption that these states have propositional contents, of the form *there is a fly around*. This claim is in tension with the standard view (accepted in Millikan, 2000) that the content of concepts is subpropositional, that is, of the form *fly*. How can states with subpropositional contents evolve? What is the relation between states with propositional content and states with subpropositional content? Millikan has extensively written about states with propositional content and states with sub-propositional content (concepts), but she has not addressed the *link* between these two kinds of states. Indeed, I will show that solving it might require departing from some of her theses.

ANTI-DESCRIPTIVISM Fourth, while I think it is clear that Millikan's views abide by CONCEPTUAL ATOMISM (but see below), it is

not completely obvious whether Millikan's view is compatible with Semantic Atomism, that is, the view that content is not even partially determined by its relation to other concepts. The key requirement that threatens to undermine Millikan's semantic atomism is the need for templates. Let me elaborate.

First of all, we saw that a conception is constituted by a template and the specific information about an entity. Now, Millikan is committed to the view that a necessary condition for a subject to learn a concept for a substance S is that it possesses an adequate template, that is, that she knows what kinds of questions can be asked concerning this substance (In 6.4.2 I will argue in more detail why she is committed to this view). According to her, a subject can be wrong about the specific information it has concerning a substance (maybe water is not tasteless, after all), but she cannot be wrong about the *kind* of questions that can be asked. For instance, if I usually track a piece of gold (say, a wedding ring), whether I am developing a concept of gold or a concept of my weeding ring depends on the template I am using. If I am disposed to project certain invariances concerning size, shape, and so on, then my concept is a concept of my wedding ring. If, instead, I am disposed to project certain invariances concerning melting point, brightness and so on, then this is surely a concept of gold (see 6.4.2). Now, that means that the content of my mental state (whether it is a concept of gold or a concept of wedding ring) depends on the template I employ. However, in order to use a certain template (the template for wedding ring), I have to use certain concepts (SIZE, SHAPE,..). A template is constituted by a set of concepts. Consequently, the content of my representation partially depends on the set of concepts I have. That contradicts Semantic Atomism. Indeed, her view should probably be classified as a particular version of Weak Semantic Descriptivism, because she holds that the referential content of a concept is partially determined by its relation to other concepts (the template).

Nevertheless, let me emphasize that, even if Millikan's account is weakly descriptivist, it greatly differs from standard descriptivist views. Whereas classical descriptivists hold that the *information* one has about a substance is what (fully or partially) determines content, the kind of concepts that according to Millikan help to determine content are the ones contained in the template rather than the specific information that fills it. In other words; the content of my concept ELEPHANT is not partially determined by my thinking it is grey, big and so on, but buy the fact that I gather information about certain kind of properties (color, size,...) rather than others. As a result, while her view should properly be classified as weakly descriptivist, the very specific role played by templates distinguishes her account from most descriptivist views in psychology (for instance, Smith and Medin, 1981) and philosophy.

In that respect, it is very likely that Millikan has been mislead by the usual conflation between what I call Semantic Atomism and Conceptual Atomism. While I think her view satisfies Weak Semantic Descriptivism, because other concepts play a fundamental role in determining content (the template), it is still a version of Conceptual Atomism. The reason is that different people can use different templates in order to identify the same substance. I can recognize cats by the number of legs and size, and a blind person can recognize them by sound and texture. The templates employed by different people can be non-overlapping. As a result, there is no particular set of non-logical concepts that a

subject must possess in order to possess a given concept and, nevertheless, the concepts I possess play a role in content determination. Thus, Millikan's account fulfills Conceptual Atomism and, at the same time, Weak Semantic Descriptivism.

Let me add that in the next chapter I will argue against the idea that templates help to determine content, so I will defend a view that can be properly be said to abide by Semantic Atomism and Conceptual Atomism .

concepts as representations Another important worry is that Millikan has not explained how the sender-receiver model that she devised for simple systems (beavers, frogs and so on) is actually instantiated at the level of concepts. The producer-consumer model that she put forward was supposed to naturalize representation and content; so, if concepts are representations, it should apply there as well. However, there has not been any systematic description of how this powerful model is supposed to work in that context. One of the tasks of next chapter is precisely to develop these ideas.

perceptual tracking Relatedly, one of the main difficulties with Millikan's account is that she fails to provide a fully naturalistic account of concepts. In particular, in her analysis of reidentifying a substance, Millikan appeals to a perceptual notion of tracking, among other things (condition 1 in Reidentification). However, we manage to perceptual track entities because perception is endowed with representational content. We track entities by representing them. Therefore, a naturalistic explanation of conceptual content based on tracking is not complete unless a naturalistic account of perceptual tracking is provided. So even if Millikan's claims are informative and specify a very particular view on the nature of concepts, she has not shown how the conceptual content can be naturalized because she relies on the act of tracking, and perceptual tracking involves representational content. As Franks and Braisby (1998, p.70) claim when discussing her account, 'Ostension (and sorting) appears to be irreducibly intentional'.

This is one of the main reasons why chapter 4 was so important. Naturalizing the representational content of perceptual states is necessary in order to provide a fully naturalistic account of concepts, because any theory of concepts will probably rely on certain perceptual abilities.

It is noteworthy that this is not a problem restricted to Millikan's theory. Similar accounts which also purport to be naturalistic suffer from exactly the same problem. Fodor (2008) for instance, has recently appealed to perception in order to provide a solution to the (horizontal) indeterminacy problem:

> We can picture you as being situated at the center of a circle that includes all but only the things you can see from here, and as being at the end of a causal chain which intersects the circumference of that circle. By assumption whatever your current perceptual thought refers to must be among the links of the chain that are in the circle. The question, to repeat, is Which such link? The best we've done so far is to reduce the number of candidates (Fodor, 2008, p. 216)
>
> Suppose that we represent the causal history of Adam's utterance by a line that runs to it passing through its referent intersecting the perceptual horizon somewhere or other (...).

> Here then is the proposal in a nutshell: imagine there is not just the actual Adam with the perspective that he actually has, but also a counterfactual Adam ('Adam2'), who is, say, three feet to the actual Adam's right. Adam2 has a (counterfactual) perspective on the (actual) visual scene; one that differs from Adam's perspective in accordance with the usual (i.e. the actual) laws of parallax. Assume that Adam2 tokens a representation of the same type that Adam does (...). The two tokens have the same referent iff Adam's line and Adam's line intersect at a link; and their referent is the link at which they intersect. (Fodor, 2008, p. 212-3)

In this quote, Fodor is suggesting that part of the answer to why my concept COW refers to cows (and not to dots in the retina or the big bang) involves (1) the fact that I am perceiving a cow but I am not perceiving the big bang and (2) the fact that if I were to be at a slightly different position I would still be perceiving a cow and not anything less distal than that.[48] However, perceptual states are intentional states, so unless a naturalistic account of perception is provided, he is just passing the buck to the perceptual domain. Since Fodor has not given any naturalistic account of the content of perceptual representations, this proposal seems to fall into the same problem.[49]

On the other hand, notice that this objection only shows that Millikan's naturalistic account of content is incomplete. In the next chapter I will argue that the way of overcoming this difficulty is by relying on the lessons of chapter 4. In the previous chapter I argued that the content of perceptual states can be naturalized and I showed how one can define a notion of tracking from the teleosemantic perspective. In defending my own teleosemantic account of concepts, I will show how the results of chapter 4 bear on an account of conceptual content.

LINGUISTIC TRACKING Finally, while I think a roughly Millikanian view on concepts can be defended if (among other things) it is supplemented with a naturalistic account of perception, there is an aspect of her theory that is in tension with the naturalization of tracking and perception: the idea that there is such a thing as linguistic tracking. Let me explain.

The point I am trying to make here is that it is extremely difficult to provide a non-intentional characterization of tracking and reidentifying in the way I required in the previous section if we accept Millikan's view on linguistic tracking. Millikan has repeatedly argued that language is a means of reidentifying substances *in exactly the same way perception is*:

> Language is just *one* medium by means of which a child perceives and hence identifies things in the world alongside a variety of other potential ways of identifying these same things. (Millikan, 1984, p. 306)
>
> Think of the matter this way. There are many ways to recognize, for example, rain. There is a way that rain

48 There are serious doubts that this account can be made to work. For instance, the following scenario is entirely plausible but would be precluded by such an account: in the actual world Adam perceives a cow and develops a concept COW, but in close possible worlds he fails to see it. For instance, he might be seeing the cow through a tiny hole in a wall.

49 At some points he seems to be relying on some Dretskian notion of information (specially in Fodor, 2008). So, he either does not have a naturalistic theory of perceptual content or he holds an account that suffers from the serious problems described in 1.2.3.

> feels when it falls on you, and a way that it looks out the window. There is a way that it sounds falling on the rooftop, "retetetetetet", and a way that it sounds falling on the ground, "shshshshsh". And falling on English speakers, here is another way it can sound: "Hey, guys, it's raining!" (Millikan, 1998, p. 64; 2000, p. 86)
>
> My claim is that having a concept grounded only through language is no different than having a concept grounded through, say, vision (Millikan, 2000, p. 90)

According to her, the only difference between acquiring a concept through language and perception is that usually (1) in the latter one can know the spatio-temporal location of the substance in relation to oneself and that (2) perception is much harder to mislead (Millikan, 2004). However, she insistingly argues that these differences do not alter the main point, which is that essentially the same process of tracking takes place in perception and language.

Now, if we follow Millikan in assuming that people are able to track and reidentify substances through language in exactly the same way people can reidentify substances by perceptual means, the task of providing a naturalistically acceptable account of the act of tracking becomes extremely difficult or impossible. I think that we cannot make sense of a naturalistic notion of tracking that encompasses at the same time perceptual tracking and what she calls 'linguistic tracking'. In other words: there seems to be no non-intentional way of describing a mechanism that is supposed to be in common between the reidentifying process that takes place in perception and in language. As a consequence, if we are to provide a fully naturalistic account of concepts by assuming a teleosemantic account of tracking, language will probably not be considered a kind of tracking.[50]

Therefore, I think that he idea that language and perception are essentially the same kind of process by means of which we track substances should be abandoned. Otherwise, a substantive and naturalistic account of tracking becomes very problematic. As I said, in the next chapter I will use the notion of perceptual tracking defined in chapter 4 (which does not apply to language) in order to provide a fully naturalistic account of concepts. Additionally, I will sketch a reasonable way in which concepts acquired by linguistic means can be accommodated (see 6.5.2).

5.2.4.4 *My strategy*

Now, I will try to improve those aspects that I think are faulty in previous teleosemantic accounts in order to provide a sufficiently detailed and plausible naturalistic account of concepts. Summing up the discussion of this chapter, there are five issues that I purport to carry out:

(1) First, I will dispense with talk of abilities. There might well be a close connection between having concepts and having certain abilities, but having a concept cannot be just identified with having a certain ability for the reasons sketched. I will assume that concepts are mental representations.

50 A related worry raised by some people is that there seems to be a crucial difference between perceptual and linguistic tracking, namely that the latter might be mediated by the intentions of agents (Gendler, 1998, p.71).

(2) I will try to describe in more detail the kind of states and structures that constitute conceptual representations by applying Third Teleosemantics to cognitive representations. Furthermore, I will address the question of compositionality, the relation between concepts and thoughts and I will show how propositional and subpropositional contents relate to each other.

(3) I will suggest an account that satisfies Conceptual Atomism and Semantic Atomism. My proposal will reject the appeal to templates in content determination, for reasons that will become clear.

(4) I will use the notion of 'tracking' defined in chapter 4 in order to spell out the process of 'reidentification' in non-intentional terms. To a great extent, that was the task of the previous chapter, but now we can see why a teleosemantic account of perception was so central for a theory of concepts. In the next chapter, I will show how the teleosemantic theory of perceptual content offered previously bears on the debate on conceptual content.

(5) Since language will not be considered another way of tracking substances, I will outline a possible way linguistic expressions (and concepts that are acquired through linguistic means) can acquire their meaning.

These are the five desiderata for the next chapter. They set the agenda for the rest of the dissertation.

5.3 Conclusions

In the first part of the chapter I have disentangled several discussions on the notion of concept and I have argued for a particular way of understanding the debate. Furthermore, I have defended a particular view on concepts: concepts are mental representations, which play particular roles in cognition. One of the tasks of the next chapter is to deepen this idea and establish certain connections with empirical data concerning perception, thought and memory.

Secondly, I have argued that none of the current naturalistic accounts of conceptual content are satisfactory. In particular, I have pointed out several difficulties with current teleosemantic theories of conceptual content, which showed that even if one accepts some kind of teleosemantic account, its application to the case of thoughts and concepts is not straightforward. Interestingly enough, we just saw that one of the most important problems with these views is that they lack a fully naturalistic account of perception and perceptual tracking. Consequently, a second major task of the next chapter is to argue how the notion of perceptual tracking defined in 4.2.3.3 can help us in the naturalization of conceptual content.

In conclusion, many questions still need to be resolved before a satisfactory naturalistic theory of concepts can be provided.

6

A NATURALISTIC THEORY OF CONCEPTUAL CONTENT

In the previous chapter I argued that there is still no satisfactory naturalistic account of conceptual content. The main goal of this last chapter is to show how the teleosemantic framework I have been developing so far can provide an original and, I think, adequate naturalistic theory of the content of concepts. In that respect, we saw that there is a wide range of issues concerning non-propositional content, the relation between thought and concepts, language and the naturalistic credentials of the account that need to be addressed. In this chapter I will attempt to suggest a solution to all these questions by drawing on the ideas discussed in chapters 4 and 5.

Now, as it was pointed out earlier, there are at least two different kinds of concepts, the ones we acquire perceptually and the ones that we acquire non-perceptually (basically, by means of composition or language). Here I will follow standard naturalistic theories in first focusing my attention on those concepts that are acquired perceptually, which some call 'perceptual concepts' (Papineau, 2006a). So, unless I say otherwise, I will be using 'concept' in order to refer to perceptual concepts. In the last section of this chapter I will outline a possible way my account can be extended to concepts acquired non-perceptually.

The chapter is organized in five main sections. First of all, I consider the complex relation between concepts and thoughts and the question whether propositional contents are more fundamental than sub-propositional contents or vice versa. In the second section, concepts are analyzed using the teleosemantic framework set up in the first part of the thesis. As we will see, since concepts require the creation of new structures in the brain, they pose several problems to Third Teleosemantics that need to be resolved. The third section is devoted to a more precise description of the brain structures that implement Third Teleosemantics in human cognition. I will develop what I think is a reasonable hypothesis: concepts are states and structures generated in memory systems. That will bring us to the Qua Problem, which will be addressed in section 4. In the last part of the chapter, I consider different ways this teleosemantic picture can account for non-perceptual concepts (i.e. concepts acquired compositionally or by linguistic means).

6.1 CONCEPTS AND THOUGHTS

There is an intimate relation between concepts and thoughts that must be specified from the beginning. This is going to be a leading thread in this chapter, so it is important to start by getting clear about this issue.

The first and obvious reason for linking the analysis of concepts to an account of thoughts is that concepts are, by definition, those mental states[1] that compose thoughts (and other propositional attitudes). Similarly, thoughts are, by definition, mental representations composed of

1 Again, I will be assuming that a part of a state of affairs is itself a state of affairs. Accordingly, since thoughts are mental states, concepts are also states.

concepts. This interdefinition explains why an account of concepts and an account of thoughts are closely entwined.

Indeed, the relationship between concepts and thoughts is not only a matter of interdefinition, but also a matter of content determination (which is certainly more problematic for us). It seems that the content of concepts is determined by its participating in contentful thoughts, while at the same time the content of thoughts seems to be determined by the content of the concepts composing them. So concepts and thoughts are interdefined and, furthermore, the direction of the determination of content seems to go in both directions. The latter feature is, I think, specially troubling. How can we get out of this circle? Should we prioritize an analysis of the content of thoughts or an analysis of the content of concepts? These are the questions I would like to address in this first section.

6.1.1 *Compositionality and Context*

At root, the problem concerning the connection between concepts and thoughts is a particular case of the more general question concerning the priority of propositional content or sub-propositional content (see below). So in this discussion I will be assuming a close connection between the debate on the relation between concepts and thoughts and the debate on the relation between sub-propositional content and propositional content.

States with propositional content are states that can be true or false, i.e. they have truth-conditions. For example, states that represent the presence of certain states of affairs, such as *there is a fly* or *there is a round black object moving in a certain way* can be assessed for truth and falsity. Thoughts are just one kind of state with propositional content. In contrast, states with subpropositional content lack-truth conditions, but they participate in more complex states with truth-conditions.[2] States that represent *fly* or *black* are simple examples. More precisely:

Subpropositional A state r has a subpropositional content iff

1. r has no truth conditions.
2. r is a constitutive part of some states with truth-conditions (i.e. propositional contents).

For instance, CHAIR has the sub-propositional content *chair* and TREE has the sub-propositional content *tree.* CHAIR and TREE have sub-propositional contents because (1) tokens of these states (alone) cannot be evaluated for truth or falsity and nevertheless (2) they are constitutive parts of more complex states that do have truth-conditions. While a token of the concept TREE is neither true nor false, a token of TREES ARE PLANTS is a (true) representation.

Note that so far in this dissertation, we have been dealing with states with propositional content. The reason is that, in general, it is very plausible that simple representational systems are more interested in the presence or absence of certain state of affairs (*the presence of food, the*

2 I will be assuming throughout the discussion that the composing elements participate in complex representations by being its syntactic parts. In other words, the notion of *participation* I am using is the one in which 'Dan' and 'tall' participate in the complex representation 'Dan is tall'. Consequently, in this discussion I will focus on the so called 'concatenative compositionality', and I will leave aside the question of 'functional compositionality' (Van Gelder, 1990).

absence of predators,..) than in the capacity for representing objects as such (see Millikan, 2000, p. 198-9). Furthermore, it is not unreasonable to suppose that the capacity to represent entities without predicating anything of them seems to require a more sophisticated mechanism (e.g. the ability to recombine certain representational states) (Sterelny, 2003). These are some of the reasons why many people think that cognitively unsophisticated organisms lack concepts (e.g. Bermudez, 2003).

So, before moving ahead and considering the complex relation between propositional and sub-propositional contents, let me illustrate with an example what would be for a simple creature to be endowed with states with sub-propositional content.

6.1.1.1 *An Example*

Consider the case of two imagined creatures, Leucippus and Xenocrates.[3] Leucippus has four brain structures, A, B, C and D. When Leucippus' mechanisms are activated (what I will represent as 'A*', 'B*', 'C*' and 'D*'), each mental state has the following contents: *There is a mouse around me now* (A*), *there is a dog around me now* (B*), *there will be a mouse around me in 30 seconds* (C*), *there will be a dog around me in 30 seconds* (D*). Let us assume that all these systems are independent, i.e. that any of these structures can be activated without any other mechanism being thereby activated.

Xenocrates has also four mechanisms 1, 2, 3, and 4, which allow him to represent the same contents as Leucippus. Nevertheless, Xenocrates' mechanisms work in a very different way. In order to represent the same contents as Leucippus, Xenocrates uses different sets of mechanisms: the content of Leucippus' mechanism A is represented by Xenocrates by means of activation in 1 and 3; the content of B by activation in 2 and 3; the content of C by activation in 1 and 4 and the content of D by activation in 2 and 4. That is, Xenocrates represents the same contents as Leucippus in the following way: *There is a mouse around me now* (1*, 3*), *there is a dog around me now* (2*, 3*), *there will be a mouse around me in 30 seconds* (1*, 4*), *there will be a dog around me in 30 seconds* (2*, 4*). In other words, the mechanisms are wired in such a way that when a mouse is present now and in 30 seconds, 1 is activated. When a dog is present now and in 30 seconds 2 is activated. When a dog or a mouse is represented as being present now, 3 is activated and when a dog or a cat is activated as being present in 30 seconds, 4 is activated.

There are many interesting questions we can raise concerning Leucippus and Xenocrates' mechanisms. For example, one striking feature of these examples is that there does not seem to be any a priori advantage of having one mechanism over the other. Both use the same number of states and can represent the same circumstances. Similarly, one might wonder whether one of the two mechanisms is easier to evolve, or which one is more likely to evolve further. However, I would like to focus on a different aspect, namely the contents attributed to these different states.

On the one hand, Leucippus seems to have the kind of representational mechanism we have been focusing on, since states A*, B*, C* and D* have propositional content. But think about Xenocrates' state 4*; this state seems to have subpropositional content, in the sense of SubPropositional. First, 4* alone can not be assessed for truth or

3 In this explanation, I follow Martinez (2010, p. 110-111).

falsity. Secondly, the activation of 4 is a *constitutive* part of states that can be assessed for truth and falsity, and indeed it contributes a very particular feature to the content of the whole state. So Xenocrates has complex states with propositional content (1* plus 2*, 2* plus 3*,...) and simple states with subpropositional content.

6.1.1.2 *Compositionality Principle and Context Principle*

Let us now move to a more controversial question: in general, what is the relation between states with propositional content and states with subpropositional content? Should we expect the propositional content of certain states to be determined by subpropositional states, or viceversa?

On the one hand, many people find it reasonable that the content of simple expressions (words, concepts,..) derives from the content of the complex expressions they participate in. This is specially plausible in the case of language; it seems that most of the time we get the meaning of a word by seeing how it is used in the context of certain sentences. In this picture, states with subpropositional content arise as constitutive parts of more complex states or structures, which carry propositional contents. This idea can be expressed more formally as follows:

Context Principle The meaning of a complex expression determines the meaning of its constituent expressions (Linnebo, 2008).

A plausible way of interpreting this principle in the context of concepts and thoughts is the following:

Context For any state r, r has a subpropositional content S in virtue of being a constitutive part of S-involving thoughts.

Context Principle entails that states with subpropositional content derive their content from states with propositional content. This idea has been explicitly endorsed by many people (though usually, for different reasons). In the context of naturalistic theories, Millikan (2004, pp. 50-21), for instance, claims:

> It is a serious mistake to suppose that the architectural or compositional meaning of a complex sign is derived by combining the prior independent meanings of its parts or aspects. Rather, the meanings of the various significant parts or aspects of signs are abstracted from the prior meanings of complete signs occurring within complete sign systems. (...) Similarly, words do not have meanings first and then get combined into sentences. Nor does the ability to think begin, say, with the ability to think 'horse', for horse and then other parts of propositions get added on later.

However, it is well known that this principle clashes with another very plausible statement that we formulated earlier, namely that the content of a thought depends on the content of the parts composing it. That is, it seems that the content of a thought like TREES ARE GREEN depends on the content of the composing concepts. This phenomenon is an example of 'compositionality' and many people think it governs thought and language. More formally, it says:

Compositionality Principle The meaning of a complex expression is determined by the constituent expressions plus the combination rules.

If we try to specify a bit more the Compositionality Principle with respect to our topic, we get the following result:

Compositionality For any thought t, t is S-involving in virtue of the fact that it is partially constituted by a state with content S.[4]

The main argument in favor of Compositionality principle is the productivity and systematicity of thought (and language) (Fodor, 19982001; Szabo, 2007). Very roughly, that thought is productive means that although we are finite beings, we can understand each of an infinitely large set of complex expressions (see 3.2.4 and 6.5.1). For instance, it is probably the first time that someone writes the sentence 'Five thousand Colombian duck hunters had dinner in the White House', and nevertheless this is a meaningful and well-formed expression that any competent English speaker should understand. On the other hand, that thought is systematic means that 'anyone who understands a complex expression e and e' build up through the syntactic operation F from constituents $e_1 ... e_n$ and $e'_1 ... e'_n$ respectively, can also understand other meaningful complex expression e'' built up through F from expressions among $e_1 ... e_n$, $e'_1 ... e'_n$ (Szabo, 2007). For example, if someone understands the expression 'John loves Mary', she also understands the sentence 'Mary loves John'. Compositionality Principle gives the resources for explaining these facts: if the content of complex expressions derives from the content of its composing parts (plus the way they are composed), we can easily account for the fact that we can understand and produce an infinite set of meaningful sentences (even if we are finite beings) and the fact that thought and language are systematic.

Now, it is fairly obvious that Context and Compositionality are incompatible. The two principles establish certain explanatory relations that go in opposite directions. Context claims that what makes a concept to have a content S is its participating in S-involving thoughts, whereas Compositionality claims that what makes a thought S-involving is its being constituted by concepts having content S. So it seems that we should give up one of them.

Fortunately, I think that there are three compatible ways of solving this paradox without (completely) rejecting Compositionality Principle and Context Principle. In the next section, I will outline the two options that maintain the idea that states with subpropositional contents derive their contents from propositional states. Afterwards, I will sketch an account that derives propositional contents from subpropositional ones.

6.1.2 *From Propositional contents to subpropositional contents*

6.1.2.1 *Weak and Strong Interpretations*

The first option in order to overcome the apparent incoherence between the two principles is to interpret Context Principle and Compositionality Principle in different ways. In general, there are two ways of understanding these claims: on the one hand, Context Principle and Compositionality Principle can be taken to establish a strong

4 I say 'A thought is S-involving in virtue of the fact that it is partially constituted by a state with content S' and not 'A thought is S-involving in virtue of the fact that it is partially constituted by a state with *sub-propositional* content S' because the state that partially constitutes the thought could have a *propositional* content as well. For instance, the sentence 'p and q', where 'p' and 'q' are sentences.

relation of content determination between complex and simple expressions. On this strong interpretation, the claim that *a being F* determines *b being G* means that *b is G* in virtue of *a being F*. Indeed, this is how we have been interpreting both claims so far, and that is the reason we derived Context from Context Principle and Compositionality from Compositionality Principle.

But there is a second way of reading these principles, if we interpret 'determination' in a much weaker sense (Linnebo, 2008; Szabó, 2007). On this view, the Compositionality Principle just says that one can read off the content of a complex expression from the contents of its composing elements, without committing itself to any view on the direction of causation or priority. On this weaker reading, Compositionality Principle merely states that one can deduce the content of the whole from the content of the parts. This is the sense in which the height of an object (plus the light source and certain laws) determines its shadow, and at the same time the shadow (plus the light source and certain laws) determines the height of an object. Similarly, if we think of a painting, one can say that its elements plus the way they are put together determines the picture and at the same time we can coherently say that the picture determines the parts and the way they are composed. Indeed, a weak interpretation of compositionality and context principles can also be found in the literature (see Fodor and Lepore, 2001; Robbins, 2005).

Of course, if we interpreted both the Context Principle and Compositionality Principle in that way, we would not be providing any substantive account of what explains (in the strong sense) that concept and thoughts have the content they indeed possess. However, we can interpret one principle in the strong sense and the other one in the weak sense. Accordingly, the idea that the content of subpropositional states derives from the content of propositional states (that is, Context principle) is compatible with the Compositionality Principle if we interpret the latter in the weak sense of determination and the Context Principle in the strong sense. According to this proposal, simple expressions have certain content in virtue of participating in complex expressions and, at the same time, it can still be true that from the content of simple expressions one can read off the content of the complex one; this is the only sense in which the content of the composing elements determines the content of complex representations.

If we take this option, the only principle that needs to be rejected is Compositionality. The reason is that this claim requires a strong interpretation of Compositionality Principle. Compositionality asserts that thoughts have certain content *in virtue of* being composed of certain contentful elements, and that follows only if the content of a complex expression is *grounded* or determined (*in the strong sense*) by the composing elements. Nevertheless, that solution would respect the intuitive force of Context Principle, Compositionality Principle, and Context.

6.1.2.2 *Partial Failure of* Compositionality principle

Secondly, even if we interpret both principles strongly, there is a way of solving the tension between them and, at the same time, keeping both Context and Compositionality. The key move is to accept that, even if Compositionality principle is true in general, in a limited set of cases it fails. Let me explore this option in some detail.

First of all, notice that most people think that Compositionality principle is probably too strong as a universal claim. For instance, it is widely accepted that there are idioms, i.e. complex expressions like *pulling one's leg* or *kicking the bucket,* whose meaning is not compositionally derived from the content of the words that compose it. The existence of idioms clashes with Compositionality Principle, because they are complex expressions whose content does not depend on the content of their composing parts.[5] So, in any event, most people would accept that Compositionality Principle is by and large true, but that it fails in certain cases.

Drawing on this occasional failures of Compositionality principle, there is a proposal that explains how the Context principle and Compositionality principle are compatible in three steps.[6] First, suppose there is a first set of complex expressions that acquire their meaning directly, without deriving their content from their parts (so, strictly speaking, they do not abide by Compositionality Principle). The organism might acquire thoughts like *the apple is on the table, the apple is on the floor, the apple is tasty* and so on. The mechanism at place that accounts for the meaning of this basic stock of meaningful states could be something like the ones described in the preceding chapters. Further, we can imagine that, by means of that process, an organism manages to produce a varied enough set of thoughts.

Secondly, the subject could work out the meanings of the parts, of which these thoughts is constituted. This process might be hard to define in detail, but for simplicity we can suppose that the contents of the states produced in the first step have largely overlapping contents (i.e. many thoughts have S-involving contents) and the organism is endowed with a certain mechanism that recognizes this coincidence. For instance, in the case depicted above, the organism might be able to identify *apple* as an element that is shared by many thoughts; similarly, in the case of Xenocrates, he might possess a mechanism that identified 4* as a proper part of many thoughts.

Finally, we only need to suppose that this basic set of concepts is used in order to derive the content of the rest of thoughts, which are produced compositionally.

In other words, we can describe a process with the following steps:

1. There is a set of thoughts T that acquire their meaning non-compositionally. The kind of process leading to these states having a certain content might be one of the several we have described in the previous chapters.

2. The meanings of the constitutive parts of thoughts in T is worked out, following Context Principle.

3. The meanings of the set of constitutive elements identified in step 2 can be used in order to create new thoughts, a stated in Compositionality Principle.

Consequently, a partial failure of Compositionality Principle (which, in any case, is very plausible) suffices for solving this problem of interdependence.

5 Of course, one could reply that they are not complex expressions, because their meaning does not derive from their composing parts, but that would render Compositionality Principle trivially true.

6 I draw this idea from Martinez (2010), who in turn took it from Garcia-Carpintero.

In conclusion, I have described two ways states with subpropositional contents may derive their content from states with propositional content, which are compatible with Context Principle and Compositionality Principle . Nonetheless, I said earlier that this is just one option in order to solve the circularity between these principles. Let me present an alternative way of resolving this issue, which I think could also be reasonably defended.

6.1.3 *From subpropositional contents to propositional contents*

In the last section, I have described two ways one could account for the subpropositional content of concepts by assuming that they derive from the propositional contents of thoughts. If one adopts this line of reasoning, one can interpret one of the principles weakly or one could hold that there is a set of basic thoughts whose content is determined directly and so compositionality fails to hold for them. However, the general teleosemantic framework I have given so far is also compatible with a different kind of solution: it could happen that the content of basic concepts is determined directly, without needing prior thoughts. How could that be possible?

As I said in the previous section, one can interpret Context Principle and Compositionality Principle in a strong or a weak sense. I previously interpreted Compositionality Principle weakly in order to make it compatible with Context principle. But one could also weakly interpret the Context Principle in order to make it compatible with Compositionality principle. In other words, one could assume that states with propositional content derive from states with subpropositional content by some process of composition, and that the Context Principle is only true to the extent that one can read off the content of the composing parts from the content of a complex expression.

Indeed, it is not unreasonable to think that this is what Millikan's (2000) and other teleosemanticists had in mind when developing their account of concepts (even if they often explicitly endorse the Context principle). According to them, concepts (understood as mental representations) are developed by organisms when they are regularly confronted with the same substance. Concepts allow subjects to reidentify the same entity at different occasions. However, when Millikan and others explain how concepts originate, thoughts do not seem to play any central role. Concepts seem to be created by direct contact with the entities they represent and their content seems to be directly determined by this perceptual relation. And, since concepts are endowed with subpropositional content, one might think that this set of naturalistic theories actually fits better with an account that takes subpropositional states to be first.

In conclusion, I have described three plausible and compatible[7] ways in which the threatening circularity between thoughts and concepts can be avoided. Having these possible strategies in mind, let us consider how concepts and thoughts can be accounted for with the tools set up in Third Teleosemantics.

7 Prima facie, the only solutions that seem to be clearly incompatible are the first and the third one, since each one interprets strongly the principle that the other interprets weakly.

Let us leave aside for a moment the connection with thoughts, and let us concentrate for a while on concepts themselves. If one takes a look at the literature, there are at least two aspects of concepts that make it difficult to see how they can be accommodated within THIRD TELEOSEMANTICS.

AMBIGUITY First of all, as I pointed out in 5.1.1.1, there is an ambiguity in the notion of 'concept'. If one considers how the notion is used, 'concept' sometimes refers to a state and sometimes to mental structures or mechanisms. As Taylor (2010, p. 84) suggests:

> Mentalese names are recurring inner representations that can be tokened again in distinct thought-episodes. Recurrent representations are constituents of beliefs. They are the things out of which structured beliefs are 'built'. The tokening of a recurring representation in a thought-episode amounts to a deployment of a concept in a thought-episode. (...) In addition to the recurring inner representations out of which though-episodes are built, there are also *standing* inner representational structures that persist across-thought episodes.[8]

In accordance with the first part of the quote, we said earlier that concepts are the constituent parts of thoughts, and thoughts are usually understood as states (my thought ARISTOTLE WAS A PHILOSOPHER is a state with the content *Aristotle was a philosopher*), so in this interpretation, concepts are states (i.e. parts of states).

In the second sense, concepts are standing structures. For instance, concepts are usually said to be possessed by subjects and stored in certain parts of the brain. However, states cannot literally be stored and it is not easy to see how a subject could be said to possess a state all his life. Rather, I think people using these expressions conceive of concepts as conceptual *structures*. Conceptual structures are mechanisms that a subject can possess, which ground a disposition to produce conceptual *states* when that is required. So we must carefully distinguish conceptual structures (brain mechanisms) and conceptual states (activations of these structures). This is a central distinction that I will try to define in more detail later, after setting up the required notions.[9]

NEW STRUCTURES There is a second central feature that needs to be addressed before concepts (and thoughts) can be properly analyzed within the teleosemantic framework I provided earlier. While the perceptual mechanisms I have focused on in chapter 4 are highly modularized and hard-wired into the brain, so that they easily satisfy SELECTION FOR and hence can have functions in accordance with ETIOLOGICAL FUNCTION, thoughts and concepts require the creation of *new structures* in the mind. That is, most conceptual structures or conceptual states we possess have not been selected for (see 5.2.3.1).

8 Let me add that Taylor (2010) does not want to identify concepts with these recurring standing structures and calls them 'conceptions' instead of 'concepts'.

9 In a way, the distinction between conceptual states and conceptual structures may account for the classic distinction between occurrent thoughts and merely dispositional thoughts. If we identify dispositional thoughts with their categorical basis, we get the same distinction between states and structures I am trying to pin down here.

Indeed, I argued that, even if they were, in general the content of a representation (and hence, of concepts) can not depend on the fact that this representation has been selected for (see 3.2.6). So, given that my thought STARS ARE BEAUTIFUL, my concept STAR or any of the structures that produce these representations have not been selected for, we cannot straightforwardly apply Third teleosemantics to them. Remember that Third Teleosemantics can accommodate new *representations*, but I have not shown yet whether it can account for the emergence of new *mechanisms*, which in turn are supposed to generate contentful representations. In fact, a possible answer to this worry (the appeal to adapted and derived functions) was already discussed and rejected in chapter 3. So this question is still more pressing for my theory.

Now, since plausibly neither conceptual states nor conceptual structures are selected for, we need to see how mechanisms can arise such that (1) they are not selected for (2) they constitute structures that produce thoughts and conceptual representations. My suggestion is that in order to account for this phenomenon, we need to focus on the general phenomenon of mental plasticity.

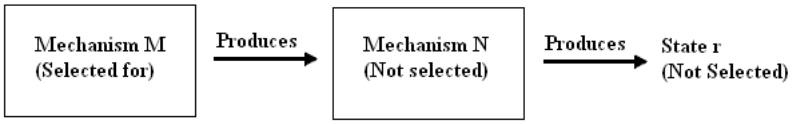
6.2.1 *Neuroplasticity*

'Plasticity' (or, more concretely, 'neuroplasticity') refers to the susceptibility to physiological changes of the nervous system, due to changes in behavior, environment, neural processes, or parts of the body. In this discussion, we are particularly interested in the process by means of which certain brain structures are designed to be modified in a certain way in order to carry out certain tasks. It is well known that the function of some parts of the brain is to generate new brain structures given certain cues. It is by producing these new structures that they manage to perform their proper activity as designed. As Kandel et al. (2000, p. 34) suggest:

> How can neural activity produce such long-term changes in the function of a set of pre-wired connections? A number of solutions for these dilemmas have been proposed. The proposal that has proven most farsighted is the *plasticity hypothesis.* (...) There is now considerable evidence for plasticity in chemical synapses.(...) Chemical synapses can be modified functionally and anatomically during development and regeneration and, most importantly, through experience and learning. Functional alterations are typically short term and involve changes in the effectiveness of existing synaptic connections. Anatomical alterations are typically long-term and consist of the growth of new synaptic connections between neurons.

A neat example of a brain structure exhibiting plasticity is the one responsible for classical conditioning. It is standardly assumed that the neuronal basis for this kind of learning is mainly located in the amygdala and, partially, the hippocampus (Eichenbaum, et al, 1999, p. 1469). Now, the function of brain structure that enables this kind of learning is to produce new structures (fundamentally, new neuronal connections) when the former brain structure is stimulated in a certain way. If we are able to understand this process and manage to fit it into

Figure 6: Schematic representation of a neuroplastic structure.



our teleosemantic schema, we will be able to accommodate structures and states that have not properly been selected for.

Let us first describe a process of neural plasticity in more abstract terms, and then we will address the question of how it applies to the particular case of thoughts and concepts.

6.2.2 *New mechanisms in Teleosemantics*

Suppose a brain contains a structure M that works in the following way; whenever a certain cue s is present, it produces a further mechanism N (say, a new set of neuronal connections), which is supposed to get activated whenever this original cue s is present. In other words, the mechanism M is such that when a certain stimulus s occurs, it produces a new mechanism N, which is supposed to produce certain state r that represents s. Crucially, mechanism N is not selected for and, nevertheless, it is supposed to produce a state following certain rules.

That is, the kind of process we need to account for is one that fulfills the following two conditions:

(C1) There is a mechanism M whose function is to produce a mechanism N such that condition C2 is fulfilled.

(C2) N is supposed to produce state r when s is present, where the relation between r and s abides by a mapping function *f*.

The schema is depicted in figure 6.

Notice that it is not obvious how C1 and C2 can be accommodated within Third teleosemantics. The definitions contained in Third teleosemantics only seem to apply to mechanisms whose function consist in producing certain *states*, rather than the production of further *mechanisms*.

In what follows, I will argue that, despite appearances, Third Teleosemantics can already account for the processes defined in C1 and C2, that is, for the emergence of new mechanisms that produce representations. Nevertheless, I will suggest to slightly modify Third Sender-Receiver in order to make the definition more transparent and manageable in the present discussion.

First of all, let us remember the last definition we gave of sender-receiver structure:

Third Sender-Receiver

> Any two systems P and C configure a sender-receiver structure if, and only if:
>
> 1. P and C have functions in accordance with Etiological Function

2. P and C have coevolved in such a way that a Normal condition for the proper performance of each system is the presence and proper functioning of the other.
3. P has two functions:
    a) The non-relational function of helping C to perform its functions.
    b) The relational function to produce a set of states R, which are supposed to map onto another set of states S in accordance with a certain mapping function *f*.
4. The function of C is to produce a set of effects E. The most proximal and most comprehensive Normal explanation for C's performance of E involves members of S.

Let us assess whether the situation described in (C1) and (C2) fits into Third Sender-Receiver. First, notice that N (the structure generated) cannot play the role of P or C in conditions 1 and 2, because it has not been selected for, and hence it does not have any function according to Etiological Function (see 2.1.2). Consequently, in order to apply Third sender-receiver we need to find two systems that have actually been selected for; the best candidates are the original system M (which creates N) and a further mechanism that consumes the state that results from the activation of N. The first important lesson, hence, is that N (the novel structure) cannot play the role of P or C in Third sender-receiver. This is a pivotal difference between new structures and the rest of representational mechanisms we have considered so far.

Accordingly, suppose that M (the mechanism that produces N) and the consumer system of N's representations satisfy condition 1 (i.e. they play the role of producer P and consumer C, respectively). Assuming that cognitive mechanisms are typically cooperating devices (which, in any event, is extremely plausible), then condition 2 is also satisfied.

Let us move to condition 3. On the one hand, 3a seems to correctly apply as it stands, because it is still true that mechanism M has as a function to help the consumer of conceptual representations to perform its functions (that seems to follow from its satisfying 2). However, 3b is more poblematic; it is not straightforward how 3b applies to the case at hand. 3b claims that the function of the producer is to generate a *state,* but in the case we are considering the relational function of mechanism M is not to produce any state, but to produce a further mechanism N (which in turn, produces a state). Should we conclude, then, than Third sender-receiver cannot be employed in the case of thoughts and conceptual representations?

There are two simple reasonings that show that this conclusion should be rejected:

- First of all, the function of mechanism M is to produce N, but N is in turn supposed to produce a given state r that belongs to a set R; so, in this case, it might still be true that the function of M is *to produce a set of states R (by means of producing a set of mechanisms).* The function of a mechanism can be to produce a state by means of producing something else (in that case, another mechanism). Consequently, 3b is straightforwardly true of mechanism M.

- Secondly, even if one of the (relational) functions of the mechanism M were to produce a mechanism N, nonetheless an additional (relational) function of the same mechanism could be *to produce a set of states R*. The fact that a mechanism has one function does not preclude its having many others.

That suggests that, strictly speaking, C1 and C2 can be accommodated in Third sender-receiver. Nonetheless, I think that for the sake of clarity and in order to facilitate the posterior discussion on systems, it is probably advisable to add an explicit appeal to the fact that a function might consist in creating a further mechanism that is supposed to perform certain tasks. Hence, in order to account for novel structures produced by functional mechanisms, I suggest to modify Third sender-receiver in the following way:

Fourth Sender-Receiver

Any two systems P and C configure a sender-receiver structure if, and only if:

1. P and C have functions in accordance with Etiological Function
2. P and C have coevolved in such a way that a Normal condition for the proper performance of each system is the presence and proper functioning of the other.
3. P has the following functions:
   a) The non-relational function of helping C to perform its functions.
   b) *In some cases, the relational function of producing a set of mechanisms N which are supposed to produce a set of states R. These states are supposed to map onto another set of states S in accordance with a certain mapping function* f.
   c) The relational function to produce a set of states R, which are supposed to map onto another set of states S in accordance with a certain mapping function *f.*
4. The function of C is to produce a set of effects E. The most proximal and most comprehensive Normal explanation for C's performance of E involves members of S mapped according to function *f*.

It is worth mentioning that this modification has no effect on Third Content or Third Representation, which remain identical. Nevertheless, again, in order to facilitate the discussion, I will call them 'Fourth Content' and 'Fourth Representation':

Fourth Representation

r is a representation iff

1. r is a member of the higher-order reproductively established family R.
2. R is a reproductively established family in virtue of being produced by a sender that satisfies Fourth Sender-Receiver.

Fourth Content

r represents s iff there are two systems p and c such that:

1. p and c are members of a Darwinian Population P and C, where P and C are systems that satisfy Fourth Sender-Receiver and Darwinian Population.
2. r is a representation (in accordance with Fourth Representation), in virtue of being produced by p.
3. s is the state that r is supposed to map onto in accordance with *f*.

Note that condition 2 of Fourth representation still holds in relation to thoughts and conceptual representations (that is, they form higher-order reproductively established families), because to produce a set of representations R is still a function of M (which plays the role of P in fourth Sender-Receiver). I will use the expression 'Fourth Teleosemantics' in order to refer to Fourth Sender-Receiver, Fourth representation and Fourth Content.

6.2.2.1 *Conceptual Representations and Conceptual Structures*

Fourth Teleosemantics is supposed to specify the conditions that any state must comply with in order to qualify as a representation. Therefore, in order to provide a specific account of *conceptual* representations we have to connect Fourth Teleosemantics with the previous issues: the distinction between propositional and sub-propositional contents, the relation between thoughts and concepts and the ambiguity between conceptual states and conceptual structures. Let us put together all the issues we have discussed so far in this chapter.

Let us start by defining conceptual structures and conceptual representations:

conceptual Structure N is r's conceptual structure iff

1. N is a mechanism that plays the role of 'N' in Fourth Sender-Receiver
2. N is supposed to produce r.
3. r is a conceptual representation, in the sense of Conceptual Representation[10]

That is, conceptual structures are mechanisms created by producers in a sender-receiver system (condition 1), which in turn are supposed to produce conceptual states (condition 2 and 3). Conceptual structures are those mechanisms that enable us to generate conceptual representations. So the next step is to define conceptual representations:

Conceptual Representation r is a conceptual representation iff

1. r is supposed to be produced by a conceptual structure N, in accordance with Conceptual Structure

10 See below.

2. r is a constitutive part of thoughts,[11] in accordance with THOUGHT.[12]
3. r has subpropositional content, in accordance with SUBPROPOSITIONAL

Let me briefly justify each condition in CONCEPTUAL REPRESENTATION. The need for condition 1 is, I hope, obvious enough. Conceptual states are produced by conceptual structures. Condition 2 is required because the rest of conditions are easily satisfied by many states and mechanisms that intuitively do not fall under the extension of 'concept'. Indeed, without 2 CONCEPTUAL STRUCTURE and CONCEPTUAL REPRESENTATION could be fulfilled by any mechanism N that simply (1) is created by a producer M (2) is supposed to produce a state r. As I said, what distinguishes conceptual representations from other sorts of representations is that they are constitutive parts of thoughts (remember that one of the problems of Millikan's view is precisely that her view fails to distinguish concepts from other representations; see 5.2.4.3). The idea that concepts are essentially tied to thoughts should be part of definition. Finally, condition 3 is included because a thought could also be a constitutive part of a thought (e.g. JOHN IS YOUNG is a constitutive part of JOHN IS YOUNG AND MARY IS SMART). In order to exclude these cases, conceptual representations have to be explicitly identified with states with subpropositional content.

Finally, since CONCEPTUAL REPRESENTATION appeals to thoughts, in order for it to be fully informative a certain clarification of thoughts needs to be provided:

THOUGHT A state t is a thought only if

1. t is a mental representation
2. t has propositional content (i.e. accuracy conditions)

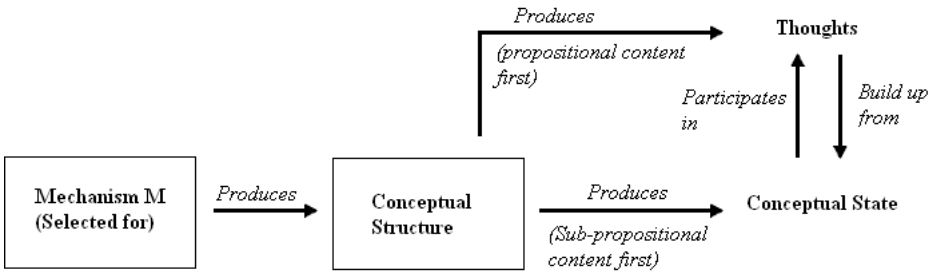
Obviously, THOUGHT only provides a set of necessary conditions for a state to be a thought, so it is not a complete definition. Unfortunately, it is not easy to provide the rest of necessary and sufficient conditions, because there are deep disagreements among philosophers. Some people think that thoughts should be defined as states with some broad functional role (Fodor, 1987), others define them as states with certain phenomenal qualities (Pitt, 2004), and many other definitions can be found. Nonetheless, for our purposes, we do not need to get into the dispute on the defining features of thoughts. It is enough if we agree on the set of states that fall under 'thought', and define concepts as the constitutive parts of these states. Fortuntately, most people agree on a large number of paradigmatic cases that should be properly called 'thoughts'. Concepts, then, can be defined as the constitutive parts of these entities, to which THOUGHT only provides a rough approximation.

Finally, let me show how this definition of concepts and thoughts connects with the discussion on mechanisms and THIRD TELEOSEMAN-

11 One question that arises here is whether a state can qualify as a concept even if it only participates in one thought. This is an issue that has been of some importance in the debate on conceptualism concerning the content of perceptual states, because if one wants to hold that in order to represent X a subject must possess a relevant concept C for specifying X, then one is also probably committed to the idea that some concepts are applied only once (for example, a concept for the very specific color hue I am perceiving right now. See Chuard, 2006). Here I will be assuming that a concept must participate in more than one thought (see Kelly, 2001), but nothing important hinges on that.

12 See below.

Figure 7: Concepts, thoughts and plasticity.



tics. The relation between conceptual structures, conceptual states and thoughts is depicted in Figrue 7.

I hope these definitions shed some light on what concepts (conceptual structures and conceptual representations) are and how they fit into the teleosemantic framework set up in previous chapters. Notice that, on this model, a mechanism N that has not been selected for can produce representations with proposotional content. Later on we will discuss which particular mechanisms instantiate this framework in the human brain.

Notice that, in this proposal, I define the function and content of conceptual structures and conceptual states without employing the notions of derived and adapted functions. Nevertheless, I appealed to mechanisms that produce other mechanisms, which in turn are supposed to produce certain states. Some readers might worry that in this theory I am smuggling in the notions I rejected in chapter 3. So before moving on I would like to discuss to what extent my account differs from Millikan's appeal to derived and adapted functions. Before moving to the brain mechanisms that actually instantiated this structures, let me shortly compare the view on concepts that follows from Fourth Teleosemantics and the account based on derived and adapted functions discussed in 3.2.6.

6.2.3 Fourth Teleosemantics *and Derived Functions.*

Since the theory I am putting forward might look similar to Millikan's theory of derived proper functions, I would like to show why the account provided in fourth Teleosemantics is different from hers and why, indeed, it avoids the objection I raised against her view.

First of all, as I set up the definitions, the mechanism M (which plays the role of P in Fourth Sender-Receiver) has (at least) two functions. On the one hand, M's function is to produce a set of mechanisms N, which are supposed to produce a set of states R. This set of states R are supposed to map onto another set of states S in accordance with a certain mapping function *f*. On the other hand, M also has the relational function of producing the set of states R (the very same states N are supposed to produce), which are supposed to map onto a set of states S in accordance with the same mapping function *f*. So M has two relational functions: one is to produce a set of mechanisms N, which in turn produce a set of states R, and the other is to produce the set of states R. So far, so good.

Now, let us consider the account of derived functions presented and rejected in 3.2.6. According to Millikan's theory, mechanisms such as N have functions, which she calls 'derived functions'. These functions have two controversial properties:

1. Derived functions are passed on from the producing device to the products.

2. Derived functions play some role in content determination.

To be more precise, we can formulate Millikan's account in the following way (let 'F_M' stand for the function of M):[13]

- $F_M[\forall x(x \rightarrow f(x))]$, and for a certain x, $f(x) = N \wedge F_N[\forall y(y \rightarrow g(y))]$

The problem we pointed out earlier is that while the function of M seems to abide by ETIOLOGICAL FUNCTION, it seems that an attribution of a function to N is not justified (i.e. claim 1 is incompatible with the general teleosemantic framework). The product of M's performing its function cannot be a mechanism that has a further function *in virtue of* being the product of M. In other words, it cannot be the case that $f(x) = N \wedge F_N[\forall y(y \rightarrow g(y))]$. If a mechanism produces another mechanism, it cannot additionally generate a function for it. In the etiological framework we are working in, we can not make sense of the idea of some functions *being carried over* to other mechanisms. Moreover, I argued that, even if these novel mechanisms had functions, they could not play any role in determining the content of the states. It is the function of mechanisms rather than the function of states what accounts for content. So claim 2 must also be rejected.

Now, one might think that my account falls prey to exactly the same problem as Millikan's, because I accept that the function of some mechanism is to produce another mechanism that is supposed to produce a state. However, there are two crucial differences between my account and the theory of derived functions. First of all, I am not assuming that the mechanism produced (N) has any function, so I reject claim 1. Secondly, I assume that once a given mechanism N is produced, the state r that N is supposed to produce is determined; similarly, r's representational content (i.e. the state s that r is supposed to map onto according to a certain mapping function *f*) is also fixed by M. The mapping function is determined by the mechanism M, which has been selected for, not by the novel mechanism N. The fact that 3c in FOURTH SENDER-RECEIVER still holds ensures this result. So, against claim 2, I am not supposing that the function of mechanism N (if it has any) plays any role in content determination. More precisely, according to my proposal:

- $F_M[\forall x(x \rightarrow f(x))]$, and for a certain x, $f(x) = N \wedge r$

That means that, in contrast to Millikan's approach: (1) produced mechanisms lack functions; (2) once a given mechanism N is produced, what this mechanism is supposed to do is also fixed; (3) even if N had any function, it would not play any role in determining the representational content of the state it produces. The function of M is to produce a further mechanism N; what this further mechanism is supposed to do

13 'f' and 'g' refer here to standard mathematical functions, not to mapping functions in the sense defined earlier.

is fully determined by M. We will see that this principle is of extreme importance when considering the tasks of conceptual structures. For the moment, it is enough if the contrast between Millikan's and my own proposal is clear.

6.3 CONCEPTS AND THE BRAIN

In the previous sections I have put forward some definitions and conceptual clarifications concerning mechanisms and states. I set up all conceptual tools that are required in order to argue that certain brain states are representations. I have shown that, if concepts are representations, then there must be certain mechanisms and functions involved in the creation of concepts, and I have also argued that this model can be easily accommodated within the general framework we have been working with.

However, I have not said yet which are the particular processes and mechanisms that determine the content of human thoughts and concepts. I have not explored the actual mechanisms and mapping functions at work in the human brain. And, arguably, it will not be possible to specify in more detail the content of concepts, unless we identify the particular mechanisms (producers, consumers,...) and their etiological functions. Furthermore, if we are able to show that this model is actually instantiated in the human brain, that will lend further support for it. In this section I would like to present some reasonable hypothesis about the actual mechanisms and functions that create and consume concepts in humans.

There are two central ideas that I would like to develop in this third section:

- The first one is that conceptual structures are produced in memory. That is, one of the functions of (some kinds of) memory is to produce conceptual structures, which in turn produce conceptual representations. Of course, the idea that concepts are representations stored in memory has been an (often implicit) assumption of most philosophers. However, I think that there are important lessons to be drawn from this fact.

- The second idea is that perceptual tracking is the fundamental ability that helps to fix the content of concepts.

Let us explicate each of these ideas, and see how they help to fill in a teleosemantic account of concepts.

6.3.1 *Concepts and Memory*

Memory is the capacity to store, retain and recover information (Klein et al, 2002, p. 439). There are different types of memory and nowadays there is a strong controversy about how should we conceive the structure and connections among these different memories. Hopefully, I think most of what I have to say about memory is quite uncontroversial, so we do not need to be committed to any polemic view on the organization of these capacities.

One of the lessons of current cognitive science is that there is not such a thing as a single and unified ability of memory. The idea that memory is a unitary process has been seriously criticized on two main grounds.

On the one hand, it has troubles accounting for dissociations, that is, cases in which one sort of memory is impaired but others work in a relatively normal fashion. On the other, it seems that some information-processing problems cannot be solved by a single mechanism, because of the different computational demands (Klein et al., 2002, p. 308). There are many different tasks memory is supposed to carry out, and it is extremely improbable that a single mechanism could fulfil them. For these reasons, it is standardly assumed that there are different parts of the brain whose function consists in storing and recovering diverse information. These memory systems differ in many aspects: the kind of information they are supposed to gather, how long the information is stored, its connection to other functional parts of the brain, etc. Many current debates concern the exact number of memory systems, and how they are related to each other and to the rest of the brain (Sternberg, 2009, ch. 5).

Cognitive scientists usually classify the different memory structures within two broad groups: declarative (or explicit) memory and non-declarative (or implicit) memory (Eichenbaum, et al, 1999; Sternberg, 2009, p. 180). Declarative memory refers to the body of memories that can be consciously recovered. It is usually divided into episodic memory (which stores specific events, like one's wedding) and semantic memory (which stores factual information, like telephone numbers). In turn, non-declarative memory is subdivided into different kinds of memory: priming, procedural memory, and so on. Sometimes, these memory systems are subdivided into several subsystems. Our present interest centers on declarative memory, which is the sort of storage that accounts for thoughts and conceptual capacities.

Now, a common thesis in philosophy and psychology is that concepts just are representations *stored* in (a certain kind of) memory, which can be recovered in working memory for certain tasks (Mahon and Caramazza, 2009). For instance, Prinz (2002) holds that concepts are mental representations stored in long-term memory. As he argues at length in his book, "(...) it would be better to say that concepts are mental representations [stored in long-term memory] that are or can be activated in working memory" (Prinz, 2002, 149). Similarly, Machery (2009) claims that this is the standard way of understanding concepts in psychology (even if he goes on to argue that this tradition is mistaken and there are no concepts). Thus, a very popular way of understanding the nature of concepts holds that concepts in humans and many other animals are nothing but mental representations that are stored in long-term memory and can be retrieved at later times.

Again, notice that, strictly speaking, representations cannot be *stored*. Representations are states of affairs, and hence either they hold or they do not hold; it does not make sense to say that a state of affairs has been stored somewhere. What is meant is that concepts are thought of as conceptual *structures* and that conceptual structures ground a disposition to produce conceptual *states* when that is required. Here, conceptual structures are identified with structures in declarative (long-term) memory, that can be recovered in working-memory. As Schachter (2001, p.33) suggests:

> Memories, according to most neurobiologists, are encoded by modifications in the strengths of connections among neurons. When we experience an event or acquire a new fact,

> complex chemical changes occur at the junctions -synapses- that connect neurons with one another.

The hypothesis, then, is that memory is the mechanism that produces conceptual structures. It is also the producer system that has been selected for and plays the role of P in FOURTH SENDER-RECEIVER.

Accordingly, the consumers of concepts are the mechanisms that use memory representations. Which are these systems? Memory is used for a wide range of activities, but two central consumers are the decision-making system (Klein et al., 2002, p. 306) and behavioral systems (Shettleworth, 2010, p. 215). The controversy around the number and functions of the different memory systems makes it extremely hard to enumerate with some precision its consumers, but they will surely involve those and other sophisticated cognitive abilities.

On the other hand, notice that, from an evolutionary point of view, developing an ability for retaining information makes a lot of sense, since it provides an important advantage over organisms lacking this ability. Memory capacities allow organisms to perform several tasks that are essential for their survival (Sherry and Schachter, 1987; Mandik, 2003). Consequently, it should not be surprising that many organisms have evolved sophisticated memory abilities. For instance, New-Caledonian Crows hide thousands of pieces of food at different places during the season of abundance, and later on (sometimes months after the hiding) they are able to remember the exact position of every one of these pieces (Shettleworth, 2010; Schachter, 2001). Moreover, not only can they remember *where* they hid a piece of food, but also *which kind* of food was it. We know that because if they discover that some kind of food is going to rot earlier than others, they recover these pieces of food before the others (Gallistell, 2010).

Another mechanism that is supposed to produce memory representations is found in lobsters. Lobsters have a complex social structure and males form a dominance hierarchy. If a group of males that have never seen each other before are put in together, they will start fighting with one another. After the fight, loosers will tend to avoid the winners, while winners will continue to display an aggressive behavior towards loosers. The way they identify other conspecifics is by recognizing certain cues in the urine of other lobsters (Breithaupt et al., 1999; Martinez, 2010). In this case, the memory representation plausibly represents particular individual lobsters, since this is what the consumer system of the representation (the avoidance mechanism) requires.

Consequently, this is the first hypothesis: conceptual structures are created in memory systems. Let us now move to the second idea: perceptual tracking is the fundamental ability in fixing conceptual content.

6.3.2 *Concepts and Perceptual Tracking*

We saw in 4.2.3.3 that the ability of perceptual tracking makes it possible for a subject to perceptually identify an entity. In particular, the entity a subject is tracking is the one that is attended by the subject, satisfies (to a certain degree) the content of the perceptual representation and the subject identifies as being the same. More precisely:

BETTER TRACKING A subject A tracks a particular entity E at $t_{1...}t_n$ iff

1. E satisfies (to a certain degree) A's perceptual content at $t_{1...}t_n$.
2. E is being attended by the subject.
3. A is disposed to behave as if the entity it is perceiving at $t_{1...}t_n$ was the same.

Here is where perceptual tracking meets a theory of concepts: the suggestion is that concepts are stored representations of those entities that a subject has been tracking during some period of time. The idea is that we usually perceptually track many entities and, in some occasions, we develop a memory structure that enables us to produce states that represent the object we have been tracking. Perceptual tracking is the ability that underlies our capacity for concept formation. As Glenberg (1997, p. 2) claims, "[most psychologists think that] the arbitrary symbols [in memory] are grounded by the perceptual system. That is, what a symbol means is what it refers to in the 'outside' world."

In chapter 4 I argued that a teleosemantic account of perceptual tracking can be satisfactorily provided (see 4.2.3.3). So, at that stage and, in contrast to other accounts, relying on perceptual content in order to fix the conceptual content should not raise any naturalistic qualms. Since the content of perceptual state has been naturalized, perception and tracking can be employed in order to naturalize conceptual content. The proposal, then, is that conceptual structures produce conceptual representations and thoughts, which are mental states that are supposed to correspond to external entities that the subject has been perceptually tracking and which are stored in memory so that they can be recovered in future occasions.

The remainder of this section is devoted to explaining these two fundamental ideas and working them out in detail.

6.3.2.1 *Tracking substances and tracking states*

Let us start by considering an important difficulty for this view, that arises from the discussion on the priority of subpropositional or propositional states. We saw that one can hold a thoughts-first or a concepts-first account. I showed that Context Principle and Compositionality Principle were compatible with the claim that propositional contents derive from subpropositional contents and also with the claim that subpropositional contents derive from propositional contents. I wanted to remain neutral concerning all these possibilities. Perhaps all these processes take place, or perhaps only some of them do. Prima facie, both the thoughts-first and the concepts-first seem plausible ways of developing higher-order representations.

Since, in principle, I see no reason for excluding the concepts-first or the thoughts-first process, I have to show how the two could be implemented in the model I am providing. Fortunately, the notion of tracking I put forward leaves room for both options. Here is the reason: I defined Better Tracking in terms of tracking *entities*, and both substances and states of affairs are entities. That is, following Better Tracking, one can track a substance (tree, water, Mama, ..) or one can track a state (the apple being on the table). So the ability of tracking enables subjects to produce structures which produce states with sub-propositional content (about trees, water or Mama) and it also enables subjects to produce structures which produce states with propositional content (about there being an apple on the table).

Let us put the idea in a different way. I said that conceptual structures are a certain kind of memory structures. But one can produce a memory of a certain substance or a memory of a certain state or event. I can remember my dog, but I can also remember the fact that my dog barked on Tuesday. So, in principle, the ability of tracking enables subjects to produce structures with subpropositional content and it also enables them as well to produce states with propositional content. Once this kind of representations are produced, the process that we described in 6.1.1 allow subjects to derive concepts or compose thoughts.

6.3.2.2 *The distality of conceptual content*

In a nutshell, the proposal that emerges from these considerations about memory and perceptual tracking is the following: generally, the function of (certain kinds of) memory is to produce mental structures that are supposed to produce thoughts or conceptual representations of entities perceptually tracked by the subject. When I am perceiving a tree, the content of my perceptual state is something like *there is an object that is brown, more or less cylindrical,...* But when I activate my memory system and produce a mental structure (which in turn will generate conceptual representations and thoughts), this state represents the substance I am tracking (say, the tree) or a given state. Memory systems (e.g. in humans, crows and lobsters) evolved because organisms needed to reidentify the very same thing under different modes and at different times (Millikan, 2000). So this kind of representations are supposed to correlate with the objects and facts themselves, and not just with the set of properties that we use in order to identify them.

Let us concentrate for a moment on the relation between the content of perceptual states and the distal content of concepts. I argued in 4.3.3 that the contents of perception are existentially quantified contents. That is, plausibly the content of my perceptual state while looking at water is something like *there is a colorless and odorless fluid.* However, I just said that when I develop a concept based on these perceptual states, the content of my concept is *water.* Why should we think that the content of my memory state is more distal than the content of the perceptual state it is based on? That is, why should we think that the content of my concept WATER is *water* rather than *colorless and odorless fluid?*

Here is where some of the lessons that we learned in chapter 4 should be recovered. There, we devised a methodological strategy in order to find out the content of a neural state: Procedure (see 4.1.4.2). In a nutshell, the idea was that we should assume that a state represents the state that most strongly activates it, *unless we have a plausible hypothesis about the needs of the consumer.* In the case of conceptual systems, however, there are already interesting and sensible hypotheses about the needs of consumer.

Recall Millikan's main argument for her view of concepts as abilities to reidentify substances (see **??**). It is extremely plausible that concepts evolved as a mechanism for gathering information about entities that can be projected and used in the future. Evolutionarily speaking, it makes a lot of sense to acquire a mechanism for remembering certain substances and states (Mandik, 2003; Shettleworth, 2010). So it is reasonable to suppose that we primarily develop memory representations in order to remember objects and events, rather than there being something with certain aspect. Millikan (2000), for instance, has extensivley

argued that the reidentification of substances is required by our inferential capacities. If we take this perspective and assume that we need concepts in order to recognize the same entities at different times and respond in the right way, then it is obvious why concepts are supposed to represent entities that go beyond perceptual content.

We can also provide a comparative argument. Think, for instance, about other organisms that have evolved a capacity for memory. In Lobsters and New-Caledonian Crows, for instance, it seems that the consumers of these memories (inferential capacities, motor systems,...) require them to identify the same *thing* or the same *fact*, rather than identifying scenarios where the same properties are instantiated. (*being brown, being elongated*,...). What inferential capacities and the relevant behavioral systems usually require for them to perform their function in the appropriate way is the presence of certain entities, rather than the instantiation of certain perceptible properties. That is, in this case it seems that the task of the memory system is often to enable the organism to reidentify the same thing in the future. So, this is a situation where the plausible needs of the system that consumes representations entail that, typically, memory representations do not represent the set of stimuli that the organism discriminates but mpore distal entities. States produced in the relevant memory systems represent things that can go beyond perceptual stimuli.

Additional support for that interpretation comes from the fact that concepts are usually activated by a great diversity of cues that come from different senses. My representation COW stored in long-term memory can be activated when I see a cow, when I touch a cow or when I hear a cow. This is important because, as many people have suggested, this variability of inputs that can elicit the very same conceptual representation is a good sign of the fact that the function of the system that generates conceptual representations is to produce a representation of something more distal than mere perceptual input (Sterelny, 2003, ch. 2; Bermudez, 2003).

Summing up the main ideas provided in this section, the proposal is schematized in figure 8.

With this picture in mind, let us provide precise definitions of conceptual content and thoughts.

6.3.2.3 *Conceptual content*

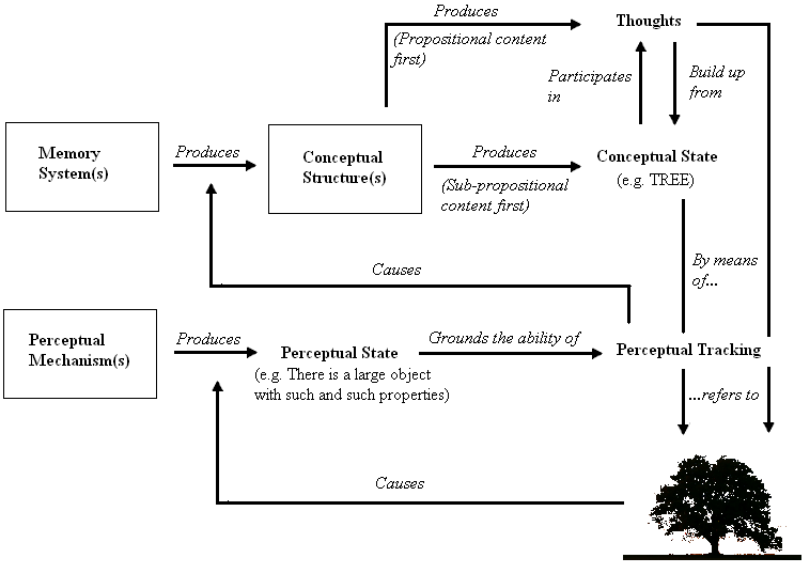
We are now in position to provide a more precise account of conceptual content. We earlier defined concepts and thoughts, but did not offer a precise description of how they acquire their content.

If one takes the concept-first strategy and thinks that the content of some concepts is determined directly (without deriving it from previous thoughts), then the following principle holds:

First Conceptual Content r is a conceptual representation of a substance E iff

1. r is a conceptual representation, in accordance with Conceptual Representation
2. r is often being employed when the subject is tracking substance E, in accordance with better Tracking.

Figure 8: A teleosemantic account of concepts.



Alternatively, if one adopts the thoughts-first strategy and holds that there is a basic stock of thoughts that comply with the first step of the process described in failure of compositionality, then:

First Thought Content t is a thought of a state E iff

1. t is a thought, in accordance with thought
2. r is often being employed when the subject is tracking state E, in accordance with better Tracking.

First Conceptual Content and First Thought Content are good first approximations, but they are not yet what we require. I have defined both principles in such a way that conceptual representations and thoughts can *only* acquire their meaning by perceptual tracking. However, it is possible that some concepts acquire their content by perceptual tracking while others derive it from the thoughts they participate in. Similarly, some thoughts may acquire their content directly, while others may derive it from the composing concepts. Therefore, we should include a third condition in each definition that allows for these alternative processes:

Second Conceptual Content r is a conceptual representation of a substance E iff

1. r is a conceptual representation, in accordance with Conceptual Representation
2. At least one of these conditions hold:
   a) r is often being employed when the subject is tracking a substance E, in accordance with better Tracking.
   b) *r derives its content from the thoughts it participates in, in accordance with* Context

Second Thought Content t is a thought of a state E iff

1. t is a thought, in accordance with thought
2. At least one of these conditions hold:
   a) t is often being employed when the subject is tracking state E, in accordance with better Tracking.
   b) *t derives its content from the concepts that compose it, in accordance with* Compositionality.

Second Conceptual Content and Second Thought Content contain many of the ideas I have been trying to develop in this chapter. In particular, they describe very clearly the connection between thoughts, concepts and perceptual tracking. They condense the main theses of the teleosemantic account of concepts I want to defend.

6.3.2.4 *Memory and Fourth Teleosemantics*

I have argued that a sender-receiver model can be found in the mechanisms responsible for the creation of concepts and thoughts. I have also shown that concepts and thoughts can be conceived as states produced in the framework described in Fourth Teleosemantics. Nevertheless, I admit that the question of the etiological functions of certain mechanisms is always difficult to settle. Consequently, in order to complete this outline of a teleosemantic theory of concepts, I would like to specify in more detail the etiological functions of the mechanisms that produce conceptual structures. In other words, I would like to show in detail how condition 3 of Fourth Sender-Receiver is satisfied in the case of conceptual structures.

So let me go back to Fourth Sender-Receiver and answer the most difficult question: what are the functions of the memory systems that produce concepts? Following condition 3 in Fourth Sender-Receiver, there are three relevant functions of the producer system (memory) to be identified:

- *The non-relational function of helping the consumer to perform its functions (condition 3a).* Memory seems to clearly satisfy the (non-relational function) of helping the alleged consumer systems (motor systems that generate behavioral responses, inferential processes,...) to perform their functions (condition 3a). I think it is extremely plausible to assume that memory systems have this kind of function; without mechanisms that used memory structures in order to carry out certain inferences or guide actions in certain ways, memory would hardly have evolved.

- *In some cases, the relational function of producing a set of mechanisms N which are supposed to produce a set of states R. These states are supposed to map onto another set of states S in accordance with a certain mapping function f (condition 3b).* A reasonable hypothesis seems to be that one of the functions of memory is to produce certain brain *structures* (e.g. neuronal connections) when certain cues are present. For instance, it is surely one of the functions of episodic memory to produce a brain structure when one is in a certain scenario. Given that John is seeing an apple on a table, the episodic memory produces a brain structure, whose activation is supposed to represent this apple being on that table. Similarly, I argued that it is reasonable to suppose that the function of

certain kinds of memory is to produce structures that generate representations of substances, rather than representations of facts. Depending on the answer we provide to this question, we will assume the propositions-first or the concepts-first solution to the circularity problem (or both). Either way, the assumption that memory has such a relational function seems extremely plausible. That suffices for justifying 3b.

- *The relational function to produce a set of states R, which are supposed to map onto another set of states S in accordance with a certain mapping function f (condition 3c).* Finally, it seems that, given what we just said, memory also has the function of producing certain mental states that are supposed to map onto certain objects, properties or facts that the subject is tracking. Indeed, the only reason some memory systems generate new brain structures is precisely for the subject to be in a position to produce representations of these objects, properties or states of affairs, so this claim is also hard to deny. This shows how 3c applies to this case.

I hope this detailed analysis helps to clarify the connection between memory structures and Fourth Teleosemantics.

6.3.3 *Some Consequences*

I think that there are numerous interesting consequences of the account I just provided, but I would like to highlight two of them: the solution to the distality problem and conceptual learning.

distality First, notice that the approach suggested here provides a solution to the distality problem of naturalistic theories. Recall that causal theories like Dretske's or Prinz's, according to which the content of a mental state is one of their causes, had the problem of having to identify which of the different objects in the causal chain was the content of the representation. What determines the fact that a concept COW refers to cows rather than cow-looking things or proximal stimuli? My account of perceptual states gives us the key for solving that problem: a concept refers to the object that satisfies the descriptive content of perceptual states. That is, if my perceptual state represents there being an object that is brown, cylindrical, 10 meters in front of me,... my concept is about the object that satisfies this description. Basically, the idea is that perceptual representations help us to pick up the object represented by conceptual states. And since a naturalistic account of perceptual content has already been provided, the whole account of conceptual content is naturalistic through and through.

learning Secondly, this proposal can account for the fact that new concepts can be learned. To learn a concept is to generate a new conceptual structure when confronted with an entity. A new conceptual structure enables the production of a new conceptual state, and hence can generate thoughts about new entities. For instance, suppose it is the first time I perceive a gun. The content of my perceptual state is something like *there is a grey and cylindrical object*. And, while I track the gun I create a conceptual structure, whose activation represents the object that satisfies this description, namely the gun.

Interestingly, I think this proposal is compatible with many current approaches to conceptual learning. Indeed, it is a way of filling in some general suggestions that are popular in the literature. For instance, Margolis (1998) has hypothesized that there must some 'sustaining mechanisms' that account for the fact that our concepts are causally linked to their referents. The process of tracking can be one of the central ways in which concepts are connected to their referents. Similarly, in a recent paper, Margolis and Laurence (2011, p. 529) suggest:

> Learning generally involves a cognitive change as a response to causal interactions with the environment. (...) learning often implicates a cognitive system that isn't just altered by the environment but, in some sense, has the function to respond as it does. For example, learning facts about the locations of various objects when entering a room isn't just a matter of having your mind altered upon perceiving the situation. The changes presumably are of the sort that our perceptual systems and related belief-fixation mechanisms are designed to subserve. (...) learning processes are ones that connect the content of an experience with the content of what is learned. The two aren't merely causally related. They are semantically related.

Analogous broad ideas are easy to find in many philosophical theories of concepts, but few of them have attempted to fill in the details of such an account. The theory I am proposing suggests a very specific hypothesis about learning.

Similarly, it is a way of filling in the idea of 'object-files' developed by Recanati (Forthcoming). The conceptual structures can be identified with the object-files that we use in order to gather information and what he calls 'epistemic-rewarding relation' can be any of the content-endowing relation we have identified so far.

Unfortunately, Second Conceptual Content still has to address an important difficulty. This remaining question is what some people call the 'qua problem'.

6.4 the qua problem

Remember the definition of Perceptual tracking provided in 4.2.3.3:

Better Tracking A subject A tracks a particular entity E at $t_{1...}t_n$ iff

1. E satisfies (to a certain degree) A's perceptual content at $t_{1...}t_n$
2. E is being attended by the subject.
3. A is disposed to behave as if the entity it is perceiving at $t_{1...}t_n$ was the same.

Now, a serious problem for the account just presented is that the relation of tracking an entity is insufficient for grounding the content of human concepts. Even if the content of my perceptual state picks up a particular entity E (say, Jack the sparrow), an entity E usually belongs to many different categories. If I am perceptually tracking a sparrow, I can be developing the concept BIRD, the concept SPARROW, the concept ANIMAL or the concept JACK THE SPARROW. So perceptually

tracking a given entity seems to be necessary but not sufficient for determining the content of the concept. This is what some people call the 'qua problem' (Devitt, 19811991; Sterelny, 1990; Prinz,2002, p. 240).

The qua problem has often been identified with the indeterminacy problem (discussed in 1.2.2.4 and 2.3.3).[14] I think that this is an important mistake. One of the main goals of chapter 2 was to show that teleosemantics can solve the indeterminacy problem in any of its formulations. I argued that teleosemantics can provide a specific content for the mental states of many organisms: for states that are automatically triggered in toads, communication signals among animals, states in perceptual processing, etc.. So, while the indeterminacy problem discussed in chapter 2 was supposed to be an endemic problem of any teleosemantic account (see Fodor, 1990), the qua problem arises only in those organisms that are able to develop many different conceptual representations while tracking a given entity. This is what happens in humans and (probably) closely related organisms, but it is not a general problem for teleosemantic accounts of content. Surely, most of the mental states of bees, toads and ants do do not suffer from the qua problem. Indeed, since I think that this problem is restricted to organisms like humans, which possess sophisticated and flexible cognitive mechanisms, the solution will also come from certain complex mechanisms possessed by cognitively sophisticated organisms.

In chapter 5 we saw that in order to solve this difficulty Millikan (1998, 2000) and Papineau (2003) appeal to the notion of template (see **??**). The strategy they both suggest is basically the same: whether I am developing a concept BIRD, SPARROW or ANIMAL depends on the kind of invariances I am disposed to project. If I am disposed to gather information about number of legs, color, size, etc.. these dispositions determine the fact that I am developing a concept SPARROW. If I am disposed to gather information about the name of the entity, where it lives, etc., I will be developing a concept of this particular sparrow. The template I am using (the kind of questions I am asking) determines the kind of substance (individual, natural kind,...) I am developing a concept of. As Papineau suggests:

> The kind of information that it is appropriate to carry from one encounter to another will vary, depending on what sort of entity is at issue. For example, if I see that some bird has a missing claw, then I should expect this to hold on other encounters with that particular bird, but not across other encounters with members of that species. By contrast, the information that the bird eats seeds is appropriately carried over to other members of the species. The point is that different sorts of information are projectible across encounters with different types of entity. If you are thinking about some metal, you can project melting point from one sample to another, but not the shape of the samples. If you are thinking about some species of shellfish, you can project shape, but not size. If you are thinking about individual humans, you can project ability to speak French, but not shirt color. And so on.

14 An illustrative example, in which the error, the indeterminacy and the qua problem are confused: "The earlier problem of error for informational theories is, in effect, a version of the qua problem" (Devitt, 1991, p. 437)

> Given this, *we can think of the referents of perceptual concepts as determined inter alia by what sort of information the subject is disposed to attach to that concept.* If the subject is disposed to attach particular-bird-appropriate information, then the concept refers to a particular bird, while if the subject is disposed to attach bird-species-appropriate information, then reference is to a species. In general, we can suppose that the concept refers to an instance of that kind to which the sort of information accumulated is appropriate. (Papineau,2006b, p. 6, emphasis added)

Hence, Millikan and Papineau suggest that whenever I am perceptually tracking an object, the content of my concept is determined by (1) the entity I am perceptually tracking (2) the template or invariances I am disposed to project. That is:

THIRD CONCEPTUAL CONTENT r is a conceptual representation of a substance E iff

1. r is a conceptual representation, in accordance with CONCEPTUAL REPRESENTATION.
2. At least one of these conditions holds:
   a) r is often being employed when the subject is tracking a substance E, in accordance with BETTER TRACKING.
   b) r derives its content from the thoughts it participates in, in accordance with CONTEXT
3. *E belongs to the category determined by the template used when developing r.*

This is the standard proposal of philosophers who attempt to provide a naturalistic approach to conceptual representations.[15] Condition 1 relies on common assumptions of naturalistic accounts of content (part 1 of this dissertation). Condition 2 relies on the idea that conceptual meaning is determined by the ability of perceptual tracking or by the thoughts it participates in, as described in chapter 4 and 6 (see 4.2.3.3). The third condition (the crucial condition in order to solve the qua problem) seeks to account for the fact that humans can develop concepts of many entities belonging to different categories by perceptually tracking a single thing E.

It is worth mentioning that, as I argued, the appeal to templates is the key feature that makes Millikan's view on concepts a version of WEAK SEMANTIC DESCRIPTIVISM (see 5.1.3). Since templates play a role in content determination, conceptual content is partially determined by the set of concepts one possesses.

But, does the appeal to templates provide a satisfactory reply to the qua problem? I have some doubts. In order to show why this proposal might be problematic, let me shortly discuss an akin solution offered for a similar problem that arose in a parallel context: phenomenal concepts.

15 See also Dickie (2010, p. 226): "The Modal Containment Principle (MCP): *a file of beliefs is 'about' an object only if the templates generated by the file's Governing Conception are templates it is possible for things of the object's category to fill.* The sense of 'possibility' relevant to the MCP is what I shall call 'categorical possibility'. This possibility is relative to a thing's category."

6.4.1 *The Qua Problem in Phenomenal Concepts*

A discussion concerning the qua problem has also arisen in the debate on phenomenal concepts. I would like to shortly discuss this topic, because there are certain lessons to be learned from this debate.

Phenomenal concepts are supposed to be a set of concepts that refer to phenomenal properties. In particular, they are some of the concepts that we employ when thinking introspectively about experiences that we are undergoing: when we think about that particular pain that we are having or that color we are experiencing. Now, one of the alleged problems of providing a full characterization of phenomenal concepts is that it is not clear what grounds the fact that a phenomenal concept refers to, say, a color type (such as *red*) rather than to a specific hue (*bright red*$_{134}$) or something intermediate (Tye, 2009a). A common reply to this objection is that a phenomenal concept refers to a kind of entity E iff the subject is disposed to use that concept when instances of E occur (Loar, 1990, Block, 2006). If I am disposed to use a concept C when confronted with red things, then it means *red*; if, instead, I am disposed to use C when confronted with bright red$_{134}$ then, it means *bright red*$_{134}$.

Now, relying on the ideas put forward in chapter 1, it is not hard to see that this crude dispositionalist view cannot be right as it stands. Nevertheless, I think it is interesting to consider why it fails.

First of all, if anything I am disposed to apply to a concept C falls under its extension, then misrepresentation is impossible (Papineau, 2006b, p. 5; see also 1.2.2.2). Misrepresentation typically occurs when I apply a concept C to an entity E but E does not fall under C's extension. However, on this crude dispositionalist view, for any entity E, if I apply C to E then E falls under the extension of C. Therefore, misrepresentation turns out to be impossible. This is surely an unacceptable result in the context of representational states such as concepts.

The second problem with this crude dispositionalist view (which I think partially explains the previous difficulty) is that it seems to take the wrong direction of explanation. It is certainly true that people tend to be disposed to apply a concept to the things that fall under its extension; but the reason they are disposed to do so is that the concept means what it means. In other words: first comes conceptual meaning, and only then the disposition to apply a concept to things under its extension, which is grounded on the first.

But, one might reply, why should we think the direction of explanation goes one way and not the other? Well, it is always difficult to argue for a certain direction of explanation. But we can apply the same reasoning that has been used at different places of this dissertation (see 1.2.4 or 5.2.2): it is standardly thought that dispositions are grounded on certain categorical properties. Which are the categorical properties that ground the disposition to apply a concept C to instances of E? A not unreasonable hypothesis is that what grounds this disposition is precisely the fact that C means E. Consequently, one cannot try to explain the fact that C means E by appealing to this disposition.

Here is an additional argument: only if the direction of explanation goes from *having certain content E* to *being disposed to apply the concept to E* can we explain the fact that sometimes a state means E and nevertheless someone is disposed to apply it to the wrong set of things. So if, as I suggest, we suppose that facts about meaning *explain* dispositions about

the use of concepts, then we can account for (1) the strong correlation between having a concept and being disposed to apply it to certain things (2) the failure of deriving meaning from this disposition.

Even Block (2006), who seems to be suggesting this crude dispositionalist view in the context of phenomenal concepts, admits this difficulty in a footnote:

> It would seem that it is because one is *taking* the image of an isosceles triangle *as a triangle-image rather than as an isosceles-triangle-image* that it functions as it does, rather than the other way around. (...). The dispositionalist view seems to get things backwards. (...). My tentative thought is that there is a form of 'taking' that does not amount to a further concept but is enough to explain the dispositions (Block, 2006 p. 39-40, emphasis added).

As Tye (2009a) notes, Block's last sentence is mysterious and has not been developed any further. So we are left with the difficulty and no apparent reply.

6.4.2 *The Qua Problem in Concepts*

Now, let us move to templates and our discussion of perceptual concepts. Millikan, Papineau and Prinz disagree with this crude dispositionalist view; they do not believe that anything I am disposed to apply a concept to falls under its extension. However, they also provide a solution based on certain dispositions of subjects. They hold that whether the concept one is developing is of an individual, of a natural kind or of any other thing depends on the kind of properties one is disposed to gather information about. Suppose John is looking at a ladybug. The proposal is that if John is disposed to gather information about the number of legs, approximate size, shape and so on he is developing a concept of ladybug; if, in contrast, he is disposed to gather information about the number of spots and its exact location, then he is developing a concept of a particular ladybug (an individual). The kind of information a subject is disposed to gather information about *determines* the category level of the referent.

Now, we might ask whether this proposal (Third Conceptual Content) suffers from the same problems as the crude dispositionalist view presented in the context of phenomenal concepts. I think that the answer is probably affirmative.

On the one hand, can this proposal based on templates account for misrepresentation? Suppose I intend to develop a concept of ladybugs in general (a natural kind), but I happen to gather the wrong *kind* of information. For instance, I might think all ladybugs have the same number of spots. Of course, this is false and, indeed, information about number of spots seems to correspond to the template of an individual (that particular ladybug). Intuitively, this is a case in which I am wrong about the template; I gather the wrong kind of information about the natural kind *being a ladybug*. However, if the template I am disposed to use determines the ontological category of the referent, then my concept is a concept about this particular ladybug (an individual). The result is that I cannot be mistaken about the template, i.e. the *kind* of information I am gathering. Thus, naturalist philosophers that appeal to templates in order to solve the qua problem are committed to the view

that one cannot be wrong about the kind of properties that are being attributed (while one can be wrong about the particular information being attributed). As a result, misrepresentation is impossible at the category level. I think that this consequence is unsatisfactory because, as a matter of fact, it seems that very often we acquire the wrong *kind* of information about new entities.

Here is another example intended to illustrate this point: suppose an explorer discovers in the Amazon a new animal that no human has ever seen before. He might intend to create a concept of this new species (i.e. of a a natural kind). However, he might have no clue as to what set of properties belong to the species and what set of properties belong to the individual. I think we have the strong intuition that, no matter what kind of information he is gathering, the explorer is developing a concept of the species; he might be completely wrong about the sort of properties he is attributing to, but the ontological category is not something that depends on the kind of information he is gathering. This is a result that theories based on templates cannot get.

Let us move to the second objection. Does the view on templates get the order of explanation right? As in the case of the crude dispositionalist, one might reasonably suppose that if a subject is disposed to gather certain kind of information about an entity, this is precisely because the concept means what it means. That is: dispositions about the kind of information that can be gathered seem to be grounded on the fact that a concept has a certain meaning and not vice versa. It is the fact that I have a concept about an individual that explains why I am disposed to infer certain kind of information and not the other way around. Again, one virtue of taking this to be the direction of explanation is that (1) it accounts for the usual correlation between having a concept and having a template (2) it can explain why I can have a concept of a natural kind or a concept of an individual and nevertheless get the template wrong. So, as in the case of the crude dispositionalist view, appealing to templates in order to solve the qua problem seems to take the wrong direction of explanation.

Therefore, I think that we need a different reply to the qua problem such that it does not rely on dispositions to gather different kinds of information.

6.4.3 *A Reply to the Qua Problem*

In contrast to standard teleosemanicists, I think that teleosemantics should offer a different kind of solution to the qua problem. Indeed, I will argue that the teleosemantic proposal I have suggested already has the resources for explaining the kind of processes that account for representations of natural kinds, individuals, and so on. Let me explain.

In this thesis (and, specially, in this last chapter) I have explained which mechanisms could give rise to conceptual representations. Using an analogous kind of mechanism, New-Caledonian Crows develop concepts of seeds (natural kind). Similarly, certain organisms develop concepts of individuals (Lobsters). These organisms have the capacity to develop mental representations of entities that belong to different ontological categories (natural kinds and individuals, respectively) because they have different mechanisms that have different functions in the sense of Etiological Function.

Thus, the idea is that an organism can develop concepts of entities that belong to different categories because it is endowed with a variety of cognitive mechanisms that have been selected for different tasks and hence are supposed to produce structures and representations of different kinds of entities. Now, if my explanation until this point has been convicing, the qua problem should not worry us: we only have to suppose that there are different brain structures that have different functions (in the sense of Etiological functions) because they have been recruited for different tasks. Some of these brain structures have been recruited for representing natural kinds, other for representing individuals, other for properties, and so on. What are these brain structures and where are they located in the brain is an empirical (and hard) question, but the mechanisms that are required for overcoming the naturalistic worry (how does certain brain activity come to represent certain things?) has been extensively described. We can attempt a sort of trascendental argument at this point: given that we have the capacity for developing concepts of entities that belong to different ontological categories (natural kinds, individuals, properties,...), our brain must be endowed with certain mechanisms (memory systems) that have been shaped by natural selection and whose function is to produce brain structures (concepts in the sense of Conceptual Structure) that produce representations of this kind of entities.

The same point can be put in a different way. Millikan, Papineau, and other philosophers have attempted to provide an answer to the qua problem by appealing to a different sort of mechanism: dispositions to collect certain kind of information. However, one of the key insights of teleosemantics is that dispositions cannot do this work. The first part of the dissertation was precisely aimed at describing an alternative set of elements that determine a representation's content, condensed in (the different versions of) Teleosemantics. So, from the teleosemantic perspective, I believe that we can offer a coherent answer by supposing that there is a set of brain structures with different etiological functions, such that each one is supposed to produce a representation of entities belonging to different kinds of substances.

Of course, whether in fact there are these brain structures is an empirical claim. But two things should be said in defense of this empirical assumption. First of all, any of the proposals on conceptual representations assume the truth of certain empirical claims. We are dealing here with facts about the human brain, so the mere fact that a proposal has empirical consequences should not be troubling. But, more importantly, I think that there is a growing body of evidence suggesting that this empirical assumption is actually true.

As I said earlier, nowadays it is standard among scientists to think that there are several memory systems, which differ (among other things) in the kind of properties they represent and how they process information (Mahon and Caramazza, 2011; Klein et al. 2002; Zahn, et al. 2007; Prinz, 2002 p., 127; Sternberg, 2009, ch. 5). There are two main sources of evidence: (1) cases of double dissociations, in which a subject has (partially or totally) lost the capacity to remember one kind of information but has a largely intact capacity for other kinds of information and (2) the fact that some of the computational tasks that a given memory system has to carry out cannot be performed by other systems (Sherry and Schachter, 1987). As a consequence, the idea that there is a single capacity for memory or even the idea that there

are few memory systems ordered sequentially has been abandoned by mainstream cognitive science. It seems to be well-established that there are different brain structures involved in the conceptualization of entities that belong to different categories and that these brain structures have been selected for performing different tasks. As Kandel et al. (2000, p.1236) argue:

> Thus, there is no general semantic memory store; semantic knowledge is not stored in a single region. Rather each time knowledge about anything is recalled, the recall is built up from distinct bits of information, each of which is stored in specialized (*dedicated*) memory stores. As a result, damage to a specific cortical area can lead to loss of specific information and therefore a fragmentation of knowledge.

Here are several examples: damage to the posterior parietal cortex results in associative visual agnosia, in which subjects can identify but cannot name objects. In contrast, damage in the occipital lobes and surrounding region can result in apperceptive visual agnosia, in which patients can name but are unable to draw objects when they are present (Farah, 2000). Similarly, visual knowledge about faces, bodies and objects is represented in different areas of the brain (Picther et al. 2009). It has also been found that certain lesions interfere with memories of living beings but not with memories of inanimate, manufactured objects. Other lesions to multimodal association areas interfere with semantic memory and others interfere with the capacity to recall episodic memories (Kandel et al. 2000, p. 1236-8). It has also been suggested that a specific brain area (the superior anterior temporal cortex) is involved with social concepts such as HONORABLE, TACTLESS, AMBITIOUS, POLITE (Zahn et al.2007). A distinction between concrete and abstract objects has also been identified (Warrington and Shallice, 1984). Similarly, different frontal regions are activated when subjects are asked to memorize different sets of stimuli (McDermott, et al., 2009). Given this body of evidence, it is not unreasonable to think that there might be different and specific areas of the brain whose function is to gather information about particular kinds of entities.

Concerning the distinction between concepts of individuals and general entities, two strong sets of evidence can be provided. On the one hand, the very distinction between semantic and episodic memory can be regarded as underpinning a distinction between the neural basis for memories of general facts (semantic memory) and memories about individual facts (Semenza, 2009, p.348; Klein et al. 2002). Secondly, it is well-established empirically that some representations about individuals are stored in different brain areas from representations about about kinds. This is the case of proper names, which are located in a different brain area from common names (probably, in the anterior temporal lobe, Semenza and Zettin, 1988; Semenza, 2009). Indeed, some scientists have suggested that this difference probably derives from the two very different tasks that representations of individuals and representations of kinds were supposed to fulfill in the evolutionary past (Klein et al. 2002; Semenza, 2009, p. 366).

These empirical data lend support to the view that the fact that humans have the capacity for representing entities that belong to different categories (individuals vs kinds, living beings vs. non-living beings,...) might be grounded on certain historical and neurological facts about

the human species. Different brain areas have different functions in the sense of ETIOLOGICAL FUNCTION and hence they have been designed to produce representations of different sorts of entities. So nothing like dispositions to project certain entities is required for concepts to acquire a determinate meaning. Therefore, I think that the qua problem can be resolved without departing from the key insights of teleosemantics.[16]

A similar view to the one defended here has recently been suggested by Lawrence and Margolis (Forthcoming):

> The first [option] relies on specialized cognitive sub-systems that are devoted to the acquisition of a given type of concept, where the acquisition systems provides a *template*[17] for concepts. The idea is that these sub-systems are activated by only certain kinds of conditions and that they fill in the template according to the ensuing experiences that the learner has. An example of this sort is the proposal that human beings have a specialized system for acquiring concepts of animals. (...) This template-based approach, with certain modifications, works well for a variety of different types of concepts apart from concepts of animals, including concepts of non-living natural kinds, concepts of individuals (name concepts), and concepts of artifacts, among others.

In conclusion, the most plausible and coherent way of solving the qua problem within the teleosemantic framework appeals to the etiological function of the mechanisms that produce representations. Different mechanisms whose function consists in producing conceptual structures (that is, 'M' in FOURTH TELEOSEMANTICS) differ in the kind of entities that they are supposed to produce a concept about. More precisely, in some mechanisms, the mathematical function f that links internal states with entities in the environment, maps onto natural kinds, while another maps onto individuals and so on. Different structures are supposed to produce representations of different sorts of entities, and that is what explains that different properties are projected.

Accordingly, I think FOURTH CONCEPTUAL CONTENT is better off than THIRD CONCEPTUAL CONTENT:

FOURTH CONCEPTUAL CONTENT r is a conceptual representation of an entity E iff

1. r is a conceptual representation, in accordance with CONCEPTUAL REPRESENTATION.
2. At least one of these conditions holds:
   a) r is often being employed when the subject is tracking a substance E, in accordance with BETTER TRACKING.
   b) r derives its content from the thoughts it participates in, in accordance with CONTEXT.

16 Moreover, notice that this solution can explain why the qua problem only affects complex cognitive systems such a the human brain, and not simpler systems such as the toad's brain.

17 Lawrence and Margolis use the word 'template' because they assume that this is essentially the same sort of solution as the one offered by Millikan and Papineau. However, I think that their account is more akin to my proposal (which differs from Millikan's and Papineau's), since it is based on the functions of certain brain structures and not on any kind of disposition.

3. *E belongs to the category determined by the function of the mechanism producing r, in accordance with* Etiological function.

In conclusion, if humans are able to represent different substances, this is because they have various mechanisms with different functions. Everything can be accounted for within the framework specified in Fourth Teleosemantics and without having to resort to dispositionalist proposals.

6.4.4 *Teleosemantics and a Theory of Concepts*

As a final remark, let me clarify the status of the view I am defending within the several classifications I made in chapter 5.

On the one hand, in contrast to Papineau and Millikan, the account I have presented can be said to abide by Semantic Atomism. Since I do not appeal to templates, the set of concepts one possesses does not play a fundamental role in content determination. Content is not determined by the information one is disposed to gather or the template one employs. Instead, we are endowed with a set of mechanisms that are supposed to track different kinds of entities. What is doing all the work in content determination is the existence of certain brain structures with certain etiological functions.

Secondly, even if Semantic Atomism is a priori compatible with many views on the structure of content, I have been following naturalistic theories in assuming Conceptual Atomism. i.e. the view that there is no set of concepts a subject must possess in order to possess any given concept (at least, any perceptual content, see below). The main reason I assume Conceptual Atomism is that the way I defend Semantic Atomism provides a plausible reply to the key objection to Conceptual Atomism. Let me explain.

In 5.1.2.1 I argued that the main difficulty of Conceptual Atomism is that this view seems to entail that one could have a concept C and have a radical misconception of C. For instance, one could have the concept BIRD and nevertheless think that birds are a piece of furniture. Now, while I think that this is indeed metaphysically possible, the conceptual atomist can explain why this scenario is extremely unlikely. If one thinks that concepts are atoms and also (partially) individuates them by referential content, then in order to have the concept BIRD one needs to track birds (in accordance with better tracking). However, it is very unlikely that one can track birds while thinking that they are a piece of furniture. While I can track birds and be wrong about many kinds of properties they have (even about its template), it is very unlikely that I can track them and think that they are a piece of furniture. Therefore, the mechanism that according to my proposal determines conceptual content helps to minimize the main problem of conceptual atomism. As a result, my way of defending Semantic Atomism makes Conceptual Atomism more plausible.

In a nutshell, the idea is that one only needs to be able to track trees in order to have the concept TREE, and no particular concept is required in order to perform this activity. In that respect, my account is similar to other atomisms (Millikan, 1998, 2000; Papineau, 2003; Margolis,1998).

That brings us to the last point: this account of conceptual content is through and through teleosemantic and naturalistic. I have not appealed to any (unanalysed) notion of perceptual tracking or disposition

that helps to fix content. In contrast, I have extensively accounted for all elements in the definition in terms of Fourth Teleosemantics and related principles. I have explained how conceptual content is determined by exclusively relying on sender-receiver systems and etiological functions. Consequently, I have offered a fully atomist and teleosemantic account of the content of perceptual concepts.

In the remainder of this chapter I will address the question of non-perceptual concepts.

6.5 derived concepts

At several points, I have distinguished perceptual concepts from concepts that we acquire by other means (e.g. language). So far, we have mainly been dealing with perceptual concepts, which are concepts that we develop in virtue of being regularly confronted with instances of a certain entity. This is why the notion of tracking was so central. For instance, I take it that for most of us, WATER, RED, TABLE or CAR are perceptual concepts. Most of this chapter has been devoted to explain how perceptual concepts are developed and how they acquire their content.

Now, since the previous debate turned only around perceptual concepts, the definition should be relativized to them:

better Conceptual Content r is a *perceptual* conceptual representation of an entity E iff

1. r is a conceptual representation, in accordance with Conceptual Representation.
2. At least one of these conditions holds:
   a) r is often being employed when the subject is tracking a substance E, in accordance with better Tracking.
   b) r derives its content from the thoughts it participates in, in accordance with Context
3. E belongs to the category determined by the function of the mechanism producing r, in accordance with Etiological function.

But there are other ways of developing concepts. Let us call 'derived concepts' all those concepts acquired by non-perceptual means. There are two main processes of non-perceptual conceptual acquisition: we can either develop them by composition or learning them by linguistic means. For instance, our concepts RED TABLE or BLACK COW are composed from simple concepts (RED, TABLE, BLACK and COW), and probably for most of us PLATYPUS or CHROMOSOME are acquired linguistically. Something must be said about these concepts.

Notice that whether a concept is perceptual or derived depends on the particular etiology of the concept. As a consequence, a concept that is derived for a subject might be perceptual for another subject.[18] Furthermore, the same concept can be perceptual and derived at the same time. I probably acquired the perceptual concept WATER when I was a child but I also have talked about it and have gathered much

18 Obviously, the distinction between perceptual, non-perceptual and derived concepts is a distinction at the level of tokens.

information about water by social means. Unfortunately, I will not be able to address in detail this complex etiology of concepts that are acquired at the same time by different media. I will only suggest certain ways in which the naturalistic account presented here can be expanded so as to account for the conceptual content of derived concepts.

6.5.1 *Composition*

First of all, it is uncontroversial that we can form complex concepts from simpler ones. For instance, my concept BROWN COW compositionally derives from my concept BROWN and my concept COW. The process of composition can be defined as that process by means of which a new concept is formed by combining two previous concepts, such that the content of the composed concept derives from the content of the composing concepts (plus the way they are composed).

Interestingly enough, notice that conceptual atomism and semantic atomism trivially fail to apply to composed concepts. On the one hand, it is obvious that I need to possess BROWN and COW in order to possess BROWN COW. So conceptual atomism fails here because some concepts are required in order to possess BROWN COW. Similarly, the content of BROWN COW is determined by the content of the composing concepts, so semantic atomism cannot be the right view of composed concepts. Nevertheless, as I said in chapter 5, these are trivial truths (e.g. see Fodor, 1998). This is the reason the debate on Conceptual Atomism and Semantic Atomism is a discussion on the structure and content of what I called 'standard' concepts. The claim that composed concepts are not atomistic is not in question. The real debate is on whether most concepts (WATER, CAR, BIRD, MAMA, etc..) are or not composed.

Now, in the discussion on Compositionality Principle I have already described some mechanisms by means of which simple representations may compose more complex ones. In that respect, since the fact that concepts compose is an uncontested fact (Fodor, 1998; Fodor and Lepore, 1992; Prinz and Clark, 2004), and given also that I do not think there is anything interesting we can add from a teleosemantic perspective, I suggest to move to the question of language.

6.5.2 *Acquiring Concepts Through Language*

Many of our concepts are acquired by means of language. So any satisfactory naturalistic account of concepts has to explain what determines the semantic properties of these concepts.

The question of language is also of central importance because an extremely popular view is social externalism, that is, the view that the content of some of our concepts derives from the content of the concepts possessed by other members of our linguistic community. Burge (1979) is well known for having designed a very compelling argument in its favor. The argument consists of three steps:

1. First, we are asked to imagine a situation where a subject, let us call it 'Jane', has many beliefs that can correctly be attributed to her with that-clauses containing 'arthritis' in oblique occurrence. That is, she thinks that she has had arthritis for four years, that it is better to have arthritis than cancer of the liver, that certain

sorts of aches are characteristic of arthritis, and so on. Crucially, she also believes that arthritis can affect her thigh. Jane does not know that arthritis is a condition of the joints only, so when she sincerely utters 'I have arthritis in my thigh' she is expressing a false belief.

2. In the second step, we imagine a counterfactual situation in which Jane's physical and phenomenal properties (nonintentionally described) and her history are held constant. However, in this scenario Jane grows up in a language community that use 'arthritis' in order to refer to a disease that also affects the thigh. The situation is counterfactual in that 'arthritis' is correctly used in this community in a sense that encompasses Jane's actual misuse.

3. The last step is a reflection on the counterfactual scenario. We are invited to judge that in the counterfactual scenario we cannot describe Jane's thoughts using 'arthritis' in oblique occurrence. The word 'arthritis' in Jane's language community does not mean *arthritis*, and we can suppose no other word does either. Since the intrinsic facts about Jane are held constant, but the beliefs are different, this is taken to show that our linguistic community plays a fundamental role in determining the content of the concepts we possess.

If that is right, the theory I have provided so far needs to be complemented with an account of socially derived concepts. I still have to explain how concepts acquired by linguistic means can be accommodated within this theory. Fortunately, I think that the tools we devised in the first part of the dissertation can also be employed in order to resolve this issue.

I think that there are various options available. The first thing to notice, however, is that acquiring concepts through language involves the semantic properties of linguistic expressions, and this is a huge field I cannot get into here. Indeed, I think that it is not unreasonable to hold that language should be treated as a kind of artifact; the functions and semantic properties of linguistic expressions might depend on the intentions of human thoughts. So here we are exploring a field that go beyond anything we have said so far.[19]

But let me present a simple way of addressing this question. Suppose John has the concept PLATYPUS and Jack lacks it, and at a certain point John tells Jack about the existence of platypus. Then, Jack develops a concept PLATYPUS that is a copy of John's concept. In the terminology we have been using here, John and Jack's concept form a (First-order) Reproductively Established Family:

First-order Ref A set of individuals $d_{1,}, d_{2,}, d_3, ..., d_n$ form a reproductively established family D iff

1. There is a set of properties $F_{1,}, F_{2,}, F_3$ such that $d_{1,}, d_{2,}, d_3, ..., d_n$ tend to instantiate a high number of these properties
2. For any d, the fact that d's ancestors had $F_{1,}, F_{2,}, F_3, ..$ in part causally explains why d has $F_{1,}, F_{2,}, F_3, ..$

19 Remember that one of the main differences of Millikan's account and the theory I am defending in this dissertation is that, according to her, human beings reidentify substances by using language in exactly the same sense we reidentify substances *in the flesh.* As I argued, I think this idea was one of the main reasons why it was very hard to spell out in more detail what constitutes the act of tracking from a Millikanian perspective (see 5.2.4.3). So in this section I am probably departing from her view.

It is not hard to see that concepts that we acquire by linguistic means are members of a First-Order Reproductively Established Family. Jack's concept has been caused by John's concept (condition 2) and will probably share many important properties (condition 1). Jack's concept *is supposed to* resemble John's in important respects. At the least, both concepts are linked to the same linguistic expression 'Platypus' and they will usually share certain information about it. Acquiring a concept by social means is a causal process that tends to produce concepts with the same properties.

Thus, since both John's and Jack's concepts belong to the same (First-order) Reproductively Established Family, Jack's concept also inherits the same meaning as John's. In other words, concepts acquired by means of language are copied from each other, and it is in virtue of this process of copy that their meaning is also carried over. That reverses the usual explanatory direction; our concepts are not the same because they have the same meaning; they have the same meaning because they belong to the same REF. The fundamental relation is being a copy of each other.

Of course, many details should be added in order to have a full theory of concepts acquired by linguistic means. The last related issue I would like to consider is how this approach can be developed in order to solve the problem of empty concepts.

6.5.3 *Empty concepts*

A recurrent objection to naturalistic theories of referential content is the existence of empty concepts. There has never existed any unicorn or any instance of phlogiston, and nevertheless UNICORN and PHLOGISTON seem to be meaningful concepts. It is not easy to see how teleosemantic accounts of content (or more generally, externalist theories; Loar, 2003) can coherently hold both theses. If these entities have never existed, they cannot have had any causal effect on our concepts, so on most naturalistic accounts UNICORN and PHLOGISTON should count as meaningless. That is clearly unsatisfactory. For one thing, it is hard to see how, if these concepts lack meaning, the following thoughts could be meaningful: PHLOGISTON DOES NOT EXIST; IF PHLOGISTON EXISTS, IT IS LIGHTER THAN OXYGEN; or JOHN BELIEVES THAT THIS METAL BAR CONTAINS PHLOGISTON. We would probably be committed to the existence of gappy contents, which, as I argued in chapter 4, are highly problematic.

A usual strategy for dealing with this sort of cases is to assume that empty concepts are composed. That is, UNICORN is in fact a concept composed of HORSE, WHITE, and so on, following the process described earlier. Accordingly, we can say that UNICORN is meaningful because the content derives from its elements that do refer, and at the same, it seems that we can assume that there has never been any unicorn. Prima facie, that looks like a satisfactory reply, but there are two main objections against this sort of proposal. First of all, this solution seems to be ad hoc. Why should we think that precisely empty concepts are composed? Why should we suppose that everyone who has the concept UNICORN or PHLOGISTON has produced this concept by composition, while the concept HORSE or TREE are not composed? Secondly, if most concepts are composed and hence have

definitions, why has conceptual analysis concerning empty concepts been so difficult? (Ryder, Unpublished).

The teleosemantic proposal sketched here provides an elegant solution to these worries. The content of the concepts I acquire through language rides piggyback on the content of the concepts from which this concept is derived. As a result, even though the concept from which I derive my concept is composed, mine can be simple (i.e. non-composed). That is: in order to solve the problem of empty concepts, we only need to assume that at some point in the causal chain, someone composed the concept UNICORN, and the rest of us have simply copied this concept (and their meaning) from him. So it might well be that all of us have a meaningful concept UNICORN and, at the same time, that none of us has composed this concept (it suffices if someone in the past did). The key feature is that we have copied this concept from other people, and that is what explains that our *empty* concept can be simple and, at the same time, meaningful. In this way, this proposal avoids having to assume that (1) anyone who has the concept UNICORN has composed this concept from simpler and referring concepts and (2) conceptual analysis of our concept UNICORN from the armchair is simple or even possible.

Again, much more should be said in order to provide a full analysis of concepts acquired by linguistic means and of empty concepts. Nevertheless, I hope I have at least suggested an interesting working hypothesis that a teleosemanticist can take in order to deal with a wide range of traditional problems concerning language and reference.

6.6 Conclusion

In conclusion, in this last chapter I have provided the main ideas of a teleosemantic account of conceptual content. I have suggested an original naturalistic theory of how conceptual content is determined, relying on the results of the previous chapters. Furthermore, the account I suggested fulfills the 5 desiderata I set up in the previous chapter:

1. Concepts are mental representations, rather than abilities

2. I described in some detail what kind of states and structures constitute concepts. I addressed the question of compositionality, the relation between concepts and thoughts and the circularity between states with propositional content and states with sub-propositional content.

3. My account abides by Semantic Atomism and Conceptual Atomism and rejects the appeal to templates in order to solve the qua problem.

4. I employ the notion of tracking defined in chapter 4 in order to naturalize the content of concepts. As a consequence, I do not leave any unanalized intentional notion in the explanans of semantic properties.

5. I explored a possible account of non-perceptual concepts.

Of course, the project of developing a detailed and complete naturalistic account of concepts would require a large discussion of many different

aspects and difficulties one can find in the literature. Nevertheless, I hope the arguments in this chapter have convincingly shown that a teleosemantic account of conceptual content is possible and, indeed, very plausible.

7

CONCLUSIONS

The goal of this dissertation was to develop a naturalistic theory of intentionality. More precisely, the project was to argue that a privileged set of semantic facts reduces to what I called 'φ-facts', that is, facts that probably metaphysically supervene on physical facts. The result, I think, is reasonably satisfactory.

First of all, we surveyed a set of theories that can be found in the literature. That was extremely useful, not only because we discovered that they were faulty, but specialy because we defined several notions that had to play an important role in the rest of the dissertation. For instance, that helped us to define the four desiderata any theory of representation has to comply with and it provided the first formulation of Relative Indication, which played an important role as a methodological strategy in chapter 4.

In the second and third chapters the Teleosemantic theory was developped. In contrast to most work done whithin this tradition, I tried to proceed slowly and justify every step in the analysis. The resultant theory is, I think, more akin to Millikanian teleosemantics than to any other naturalistic theory. Nevertheless, some disagreements where pointed out and carefully argued. With this framework in mind, some classical and recent objections were addressed and replied.

I the second part, I showed how the model meticulously spelled out in the first three chapters of this dissertation is actually instantiated in the human brain. Chapter 4 showed that the computational structure of perceptual mechanisms can actually be regarded as set of sender-receiver systems. That enabled us to naturalize the content of perceptual experiences, as well as to provide a foundational theory for representational talk in neuroscience.

The last two chapters were devoted to concepts. Since the literature on concepts is specially messy, I decided to discuss in some detail the different issues at stake and various approaches. Once the field was clarified, I presented the problems of extant naturalistic theories of content, and in the last chapter I provided my own account. On the view I defend, concepts are mental representations produced by brain structures generated in memory that we employ in thought, which allow us to track substances at different occasions and through different media. I explained what kind of mechanisms must be in place in order to porduce contentful states and structures that are not selected for.

Summing up, I think I have presented a plausible account of how an important set of intentional states can be naturalized. Furthermore, this theory opens to door to a whole range of questions that call for future research. On the one hand, there are many question that remained unanswered: Can this framework be used in order to naturalize imperative content? Can the existence of loops and top-down influences in the brain be accommodated within the sender-receiver structure? Is a teleosemantic theory compatible with a non-atomistic view of concepts? These are some of the questions that should be explored in detail.

In that respect, I think there are two lines of research that deserve special attention. On the one hand, we saw that neural systems constitute

a huge field in which sender-receiver models can be identified at many different levels: at the level of chemistry, single neurons, networks, and so on. Consequently, one important line of research would explore further the prospects of a neuroteleosemantic account. A second largely unexplored field is to investigate how, once this privileged set of mental facts is naturalized, the rest of semantic states can also be reduced to φ- facts. For instance, there is the very interesting question of human language, which I scarcely addressed in the final chapter, but also the question of artifacts.

As a final conclusion, then, I think the original hypothesis, according to which the semantic properties of a privileged set of semantic facts can be naturalized, has been satisfactorily defended. This conclusion opens a whole range of new and challenging questions and suggests a renewed look at old issues. Future research is required.

Part III

APPENDIX

APPENDIX: DEFINITIONS

A.1 NATURALISM

LOCAL φ−PHYSICALISM Semantic facts metaphysically supervene on (indeed, reduce to) φ- facts.

A.1.1 *Naturalistic Theories*

RESEMBLANCE A state R represents S iff R resembles S.

CRUDE CAUSAL ACCOUNT R represents S iff R was caused by S.

STRONG INDICATION Structure R has the fact that t is F as its semantic content = R carries the information that t is F in digital form

WEAK INDICATION A state R represents state S iff $P(S \mid R) > P(S)$

RELATIVE INDICATION 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 than are members of any other natural kind.

ASYMMETRY R represents S iff:

1. *S cause R*'s is a law
2. For all Ts that are not Ss, if Ts actually cause Rs, then the Ts causing Rs is asymmetrically dependent on the Ss causing Rs

A.1.2 *Desiderata*

Error Problem: A semantic theory suffers form the Error Problem if it does not allow for cases of misrepresentation

Adequacy Problem: A theory suffers from the adequacy problem if the content it warrants greatly and systematically diverges from the content warranted by science and common sense.

Indeterminacy Problem A theory suffers from the indeterminacy problem if it warrants multiple content attributions in cases where science and common sense warrant a single content.

Normativity A metasemantic theory suffers from the normativity problem if it cannot account for the normative difference between cases of successful representation and cases of misrepresentation.

A.2 REPRODUCTIVELY ESTABLISHED FAMILIES

REPRODUCTIVELY ESTABLISHED FAMILY A group of individuals $d_1, d_2, d_3, \ldots d_n$ form a reproductively established family D iff $d_1, d_2, d_3, \ldots d_n$ tend to resemble each other in important ways because they are the result of some causal process of copy.

First-order Ref A set of individuals $d_{1,}, d_{2,}, d_3, ..., d_n$ form a reproductively established family D iff

1. There is a set of properties $F_{1,}, F_{2,}, F_3$ such that $d_{1,}, d_{2,}, d_3, ..., d_n$ tend to instantiate a high number of these properties
2. For any d, the fact that d's ancestors had $F_{1,}, F_{2,}, F_3, ..$ in part causally explains why d has $F_{1,}, F_{2,}, F_3, ..$

Higher-order Ref A set of individuals $d_{1,}d_{2,}d_{3,...}d_n$ form a higher-order reproductively established family D iff it is a function of a device that belongs to a *first-order* reproductively established family to produce them.

A.3 function, selection and darwian populations

Darwinian Population

D forms a Darwinian Population only if the following conditions are met:

(a) *Replication*: Members of D must form a reproductively establish family, in accordance with Reproductively established family

(b) *Variation*: The replication of members of D included some changes in some of its members.

(c) *Environmental interaction*: The interaction of members of D with certain environmental circumstances determined differential replication among its members.

Selection For

D is *selected for* F iff:

1. D forms a Darwinian Population, in accordance with Darwinian Population
2. F is an effect of some members of D
3. F is the effect that (in a preponderant number of cases) causally explains why differential replication favored certain members of D that could do F.

Etiological Function

A trait d has the function F iff:

1. d is a member of D.
2. D forms a Darwinian Population, in accordance with Darwinian Population
3. D has (recently) been selected for performing F, in accordance with Selection For

A.4 teleosemantics

A.4.1 *First Teleosemantics*

First Sender-Receiver

Any two systems P and C configure a sender-receiver structure if, and only if:

1. P and C have functions in accordance with Etiological Function
2. P and C have coevolved in such a way that a Normal condition for the proper performance of each is the presence and proper functioning of the other.
3. P has two functions:
    a) The non-relational function of helping C to perform its functions.
    b) The relational function of producing state R when another state of affairs S obtains.
4. The function of C is to produce an effect E. The least detailed and most comprehensive Normal explanation for C's performance of E involves S.

First Representation R is a representation iff R is a state produced by a sender P, which satisfies with First Sender-receiver.

First Content

R represents S iff there are two systems P and C such that:

1. P and C configure a sender-receiver structure, in accordance with First Sender-Receiver.
2. R is a representation, in accordance with First Representation.
3. The most proximal and most comprehensive Normal explanation for C's performance of its functions when R obtains involves S.

A.4.2 *Second Teleosemantics*

Second Sender-Receiver

Any two systems P and C configure a sender-receiver structure if, and only if:

1. P and C have functions in accordance with Etiological Function
2. P and C have coevolved in such a way that a Normal condition for the proper performance of each is the presence and proper functioning of the other.
3. P has two functions:
    a) The non-relational function of helping C to perform its functions.
    b) The relational function of producing state R when another state of affairs S obtains.
4. The function of C is to produce an effect E. The least detailed and most comprehensive Normal explanation for C's performance of E involves S.

Second Representation

r is a representation iff

1. r is a member of the reproductively established family R.

2. R is a reproductively established family of states produced by a sender that satisfies Sender-Receiver.

Second Content

r represents s iff there are two systems p and c such that:

1. p and c are members of the Darwinian populations P and C, where P and C are systems that satisfy Sender-Receiver
2. r is a representation (in accordance to Representation) in virtue of being produced by p.
3. The least detailed and most comprehensive Normal explanation for c's performance of its functions when members of R obtain involves some s, which are members of (a REF-type or Instance-Type) S.

A.4.3 *Third Teleosemantics*

Third Sender-Receiver

Any two systems P and C configure a sender-receiver structure if, and only if:

1. P and C have functions in accordance with Etiological Function
2. P and C have coevolved in such a way that a Normal condition for the proper performance of each system is the presence and proper functioning of the other.
3. P has two functions:
   a) The non-relational function of helping C to perform its functions.
   b) The relational function to produce a set of states R, which are supposed to map onto another set of states S in accordance with a certain mapping function *f*.
4. The function of C is to produce an effect (or set of effects) E. The least detailed and most comprehensive Normal explanation for C's performance of E involves members of S.

Third Representation

r is a representation iff

1. r is a member of the higher-order reproductively established family R.
2. R is a reproductively established family in virtue of being produced by a sender that satisfies Sender-Receiver.

Third Content

r represents s iff there are two systems p and c such that:

1. p and c are members of a Darwinian Population P and C, where P and C are systems that satisfy First Sender-Receiver and Darwinian Population.

2. r is a representation (in accordance with Third Representation), in virtue of being produced by p.
3. s is the state that r is supposed to map onto in accordance with *f*.

A.4.4 *Fourth Teleosemantics*

Fourth Sender-Receiver

Any two systems P and C configure a sender-receiver structure if, and only if:

1. P and C have functions in accordance with Etiological Function
2. P and C have coevolved in such a way that a Normal condition for the proper performance of each system is the presence and proper functioning of the other.
3. P has the following functions:
   a) The non-relational function of helping C to perform its functions.
   b) In some cases, the relational function of producing a set of mechanisms N which are supposed to produce a set of states R. These states are supposed to map onto another set of states S in accordance with a certain mapping function *f*.
   c) The relational function to produce a set of states R, which are supposed to map onto another set of states S in accordance with a certain mapping function *f*.
4. The function of C is to produce a set of effects E. The most proximal and most comprehensive Normal explanation for C's performance of E involves members of S mapped according to function *f*.

Fourth Representation

r is a representation iff

1. r is a member of the higher-order reproductively established family R.
2. R is a reproductively established family in virtue of being produced by a sender that satisfies Fourth Sender-Receiver.

Fourth Content

r represents s iff there are two systems p and c such that:

1. p and c are members of a Darwinian Population P and C, where P and C are systems that satisfy Fourth Sender-Receiver and Darwinian Population.
2. r is a representation (in accordance with Fourth Representation), in virtue of being produced by p.
3. s is the state that r is supposed to map onto in accordance with *f*.

A.5 METHODOLOGICAL PRINCIPLE

RELATIVE INDICATION*

As a working hypothesis, assume that R has as its extension the members of Q if and only if:

1. It can plausibly be assumed that members of Q were present in Normal circumstances.
2. Members of Q are more efficient in their causing of R than are members of any other kind.

PROCEDURE

1. First, consider how the producer system generates a representation. In particular, find out which stimulus Q most strongly elicits a given mental state, in accordance with RELATIVE INDICATION*. As a first hypothesis, suppose that this system represents stimulus Q.
2. Secondly, find out whether it is plausible to hold that this state of affairs is what the consumer system needs in order to perform its own functions in a Normal way, as stated in THIRD TELEOSEMANTICS. The latter is what really determines content, but since the needs of the consumer system are often hard to assess, the best working hypothesis we have when addressing complex systems is that a particular brain structure represents whatever it is sensitive to. Once we know what a system most strongly reacts to, we should consider whether it is reasonable to hold that this state of affairs is what is Normally needed for the consumer to perform its functions in a Normal way.
3. Finally, if we have good reasons for thinking this state is not what the consumer system needs, then we will have to reconsider the content of the representation in light of the needs of the consumer-system. The motto is the following: *in case of disagreement, the needs of the consumer system prevail.* That means that, in some situations, what a producer system is sensitive to might not qualify as the represented state of affairs because the needs of the consumer system are different. Nonetheless, I pointed out some reasons for thinking that the methodological strategy will probably be useful because, very often, the state that satifies RELATIVE INDICATION* will be the state that satisfies THIRD TELEOSEMANTICS.

A.6 DEBATE ON CONCEPTS

CONCEPTUAL ATOMISM For any standard concept C, there is no particular set of non-logical concepts S, such that a subject needs to posses S in order to possess C.

Conceptual Structuralism For any standard concept C, there is a particular set of non-logical concepts S, such that a subject needs to posses S in order to possess C.

Classical Theory For any standard concept C, there is a particular set of non-logical concepts S, such that a necessary and sufficient condition for a subject to possess C is that it possesses S. All and only members of S are involved in the definition of C

Non-Classical Theory For any standard concept C, there is a particular set of non-logical concepts (or beliefs) S, such that possessing a sufficient number of concepts (or beliefs) of S is a necessary and sufficient condition for a subject to possess C.

Semantic Atomism The *referential* content of a concept is not determined by its relation to other concepts.

Strong Semantic Descriptivism The referential content of a concept is fully determined by its relation to other concepts.

Weak Semantic Descriptivism The referential content of a concept is partially determined by its relation to other concepts.

A.7 tracking, concepts and thoughts

Subpropositional A state r has a subpropositional content iff

1. r has no truth conditions.
2. r is a constitutive part of some states with truth-conditions (i.e. propositional contents).

Context Principle The meaning of a complex expression determines the meaning of its constituent expressions.

Context For any state r, r has a sub-propositional content S in virtue of being a constitutive part of S-involving thoughts.

Compositionality Principle The meaning of a complex expression is determined by the constituent expressions plus the combination rules.

Compositionality For any thought t, t is S-involving in virtue of the fact that its partially constituted by a state with content S.

Better Tracking A subject A tracks a particular entity E at $t_{1...}t_n$ iff

1. E satisfies (to a certain degree) A's perceptual content at $t_{1...}t_n$
2. E is being attended by the subject.
3. A is disposed to behave as if the entity it is perceiving at $t_{1...}t_n$ was the same

conceptual Structure N is r's conceptual structure iff

1. N is a mechanism that plays the role of 'N' in Fourth Sender-Receiver

2. N is supposed to produce r.
3. r is a conceptual representation, in the sense of Conceptual Representation

Conceptual Representation r is a conceptual representation iff

1. There is a conceptual structure N, in accordance with Conceptual Structure
2. r is supposed to be produced by N
3. r is a constitutive part of thoughts, in accordance with Thought

Thought A state t is a thought only if

1. t is a mental representation
2. t has propositional content (i.e. accuracy conditions)

Second Conceptual Content r is a conceptual representation of a substance E iff

1. r is a conceptual representation, in accordance with Conceptual Representation
2. At least one of these conditions hold:
   a) r is often being employed when the subject is tracking a substance E, in accordance with better Tracking.
   b) *r derives its content from the thoughts it participates in, in accordance with* Context

Second Thought Content t is a thought of a state E iff

1. t is a thought, in accordance with thought
2. At least one of these conditions hold:
   a) t is often being employed when the subject is tracking state E, in accordance with better Tracking.
   b) *t derives its content from the concepts that compose it, in accordance with* Compositionality.

BIBLIOGRAPHY

M. Abrams. Teleosemantics without natural selection. *Biology and Philosophy*, 20(1):97–116, 2005. (Cited on page 134.)

F. Adams and K Aizawa. Causal theories of mental content. *Stanford Encyclopedia of Philosophy*, 2010. (Cited on pages 38, 42, and 43.)

N. Agar. What Do Frogs Really Believe? *Australasian Journal of Philosophy*, 71(1):1–12, 1993. (Cited on page 165.)

K. Akins. Of sensory systems and the 'aboutness' of mental states. *The Journal of Philosophy*, 93(7):337–372, 1996. (Cited on pages 188 and 190.)

C. Allen. Animal concepts. *Behavioral and Brain Sciences*, 21(1):66–66, 1998. (Cited on page 250.)

L. Anton. Equal rights for swamppersons. *Mind and Language*, 11(1): 70–75, 1996. (Cited on page 155.)

D. Armstrong. *A World of States of Affairs*. Cambridge Studies in Philosophy, Cambridge, 1997. (Cited on pages 22 and 79.)

D. Armstrong. *Truth and Truthmakers*. Cambridge University Press, 2004. (Cited on page 79.)

M. Artiga. Learning and selection processes. *Theoria*, 25:197–209, number = 2,, 2010. (Cited on page 72.)

M. Artiga. Re-organizing organizational accounts of function. *Topoi*, 6: 105–124, 2011. (Cited on pages 51, 53, and 64.)

F. Ayala. Teleological Explanations in Evolutionary Biology. *Philosophy of Science*, 37(1):1–15, 1970. (Cited on page 51.)

K. Bach. Searle against the world: How can experiences find their objects? In S. L. Tsohatzidis, editor, *John Searle's Philosophy of Language: Force, Meaning, and Mind*. Cambridge University Press, 2007. (Cited on page 204.)

G. Bealer. A theory of concepts and concepts possession. *Philosophical Issues*, 9:261–301, 1998. (Cited on page 219.)

W. Bechtel. Representations and cognitive explanations: Assessing the dynamicist challenge in cognitive science. *Cognitive Science*, 22(3): 295–317, 1998. (Cited on pages 164 and 189.)

W. Bechtel. Representations: Freom neural systems to cognitive systems. In J. Mundale W. Bechtel, P. Mandik and R. S. Stufflebeam, editors, *Philosophy and the Neurosciences: A Reader*, pages 332–349. OxfordUniversity Press, 2000. (Cited on page 178.)

J. L. Bermudez. *Thinking Without Words*. Oxford University Press, 2003. (Cited on pages 169, 259, and 279.)

B. Bertenthal. Origins and early development of perception, action and representation. *Annual Review of Psychology*, 47:431–459, 1996. (Cited on pages 164 and 193.)

N. Block. Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, 10:615–678, 1986. (Cited on page 214.)

N. Block. On a confusion about a function of consciousness. In N. Block; O. Ianagan; G. GÃŒzeldere, editor, *The Nature of Consciousness*, pages 375–417. MIT Press, 1997a. (Cited on page 165.)

N. Block. Conceptual role semantics. In Edward Craig, editor, *Routledge Encyclopedia of Philosophy*. Routledge, 1997b. (Cited on page 26.)

N. Block. Max black's objection to mind-body identity. *Oxford Review of Metaphysics*, 2006. (Cited on pages 286 and 287.)

Ch. Boorse. Wright on functions. *The Philosophical Review*, 85(1):70–86, 1976. (Cited on pages 52 and 60.)

R. Boyd. Homeostasis, species, and higher taxa. In R. Wilson, editor, *Species: New Interdisciplinary Essays*, pages 141–185. MIT Press, 1999a. (Cited on pages 96, 144, and 156.)

R. Boyd. Kinds, complexity and multiple realization: comments on millikan's historical kinds and the special sciences. *Philosophical Studies*, 95:67–98, 1999b. (Cited on pages 96 and 144.)

T. Breithaupt, D. P. Lindstrom, and J. Atema. Urine release in freely moving catheterised loobsters (homarus americanus) with reference to feeding and social activities. *The Journal of Experimental Biology*, 202:837–844, 1999. (Cited on page 276.)

B. Brewer. Perception and content. *European Journal of Philosophy*, 14(2): 165–181, 2006. (Cited on page 205.)

T. Burge. Individualism and the mental. *Midwest Studies in Philosophy*, 14:73–121, 1979. (Cited on page 294.)

T. Burge. *Foundations of Mind*. Oxford University Press, 2007. (Cited on pages 155 and 211.)

T. Burge. *The Origins of Objectivity*. Oxford University Press, 2010. (Cited on pages 33, 37, 78, 80, 131, 165, 167, 168, 169, 172, 173, 202, 204, and 211.)

J. Campbell. *Reference and Consciousness*. Oxford University Press, 2002. (Cited on pages 154, 202, and 205.)

J. Canfield. Teleological explanations in biologys. *The British Journal for the Philosophy of Science*, 14:285–295, 1964. (Cited on page 60.)

R. Cao. A teleosemantic approach to information in the brain. *Biology and Philosophy*, 27:49–71, 2012. (Cited on pages 73, 141, 165, 166, 167, 170, and 182.)

S. Carey. *The Origin of Concepts*. Oxford University Press, 2009. (Cited on pages 210, 213, 218, and 228.)

G. Carlson. Names, and what they are names of. *Behavioral and Brain Sciences*, 21:69–70, 1998. (Cited on page 248.)

P. Carruthers. *Phenomenal Consciousness: a Naturalistic Theory*. Cambridge University Press, 2000. (Cited on page 193.)

P. Carruthers. *The Architecture of the Mind: massive modularity and the flexibility of thought*. Oxford University Press, Oxford, 2006. (Cited on page 210.)

A. Chakravartty. Scientific realism. *The Stanford Encyclopedia of Philosophy*, 2011. (Cited on page 23.)

D. Chalmers. *The Conscious Mind: In search of a Fundamental Theory*. Oxford University Press, 1996. (Cited on page 22.)

D. Chalmers. Perception and the fall from eden. In T. S. Gendler and J. Hawthorne, editors, *Perceptual Experience*. Oxford University Press, 2006. (Cited on page 202.)

N. Chomsky. Review of verbal behavior by B.F. Skinner. *Language*, 35 (1):26–57, 1959. (Cited on page 72.)

W. D. Christensen and M. H. Bickhard. The process dynamics of normative function. *The Monist*, 12(6):795–823, 2002. (Cited on page 63.)

P. Chuard. Demonstrative concepts without reidentification. *Philosophical Studiest*, 130(2):153–201, 2006. (Cited on page 271.)

A. Clark and J. Prinz. Putting concepts to work: Some thoughts for the twenty-first century. *Mind and Language*, 19(1):57–69, 2004. (Cited on pages 218, 223, and 294.)

C. Comer and V. Leung. The vigilance of the hunted: Mechanosensory-visual integration in insect prey. In F. R. Prete, editor, *Complex Worlds from Simpler Nervous Systems*, pages 313–335. MIT Press, 2004. (Cited on page 182.)

D. Croll; CH. Clark; A. Acevedo; B. R. Tershy; S. Flores; J. Gedamke; J. Urban. Only male fin whales sing loudsongs. *Nature*, 417:809, 2002. (Cited on page 129.)

R. Cummins. Functional analysis. *Journal of Philosophy*, 72:741–765, 1975. (Cited on page 60.)

R. Cummins. *Meaning and Mental Representation*. The MIT Press, 1989. (Cited on pages 26, 27, and 44.)

M. Owren M. Ryan D., Rendall. What do animal signals mean? *Anmal Behaviour*, 78:233–240, 2009. (Cited on page 138.)

D. Davidson. Knowing one's own mind. In *Proceedings and Addresses of the American Philosophical Association*. American Philosophical Association, 1987. (Cited on page 149.)

M. Davies. Perceptual content and local supervenience. *Proceedings of the Aristotelian Society*, 66(1):1–45, 1992. (Cited on page 204.)

R. Dawkins. *The Selfish Gene*. Oxford University Press, 1976. (Cited on page 138.)

K. de Queiroz. The general lineage concep oft species and the defining properties of the species category. In R. A. Wilson, editor, *Species: New Interdisciplinary Essays*. MIT Press, 1999. (Cited on pages 56 and 156.)

C. DeCharms and A. Zador. Neural representation and the cortical code. *Annual Review of Neuroscience*, 23(1):613–647, 2000. (Cited on pages 174 and 177.)

C. Delancey. Ontology and teleofunctions: A defense and revision of the systematic account of teleological explanation. *Synthese*, 150(1): 69–98, 2006. (Cited on page 65.)

D. Dennett. *Kinds of Minds*. Basic Books, 1996. (Cited on page 62.)

M. Devitt. *Designation*. Columbia University Press, 1981. (Cited on page 284.)

M. Devitt. Naturalistic representation. *British Journal for the Philosophy of Science*, 42(3):425–443, 1991. (Cited on pages 241 and 284.)

D. Dickie. We are acquainted with ordinary things. In *New Essays on Singular Thought*. Oxford University Press, 2010. (Cited on pages 198 and 285.)

F. Dretske. *Knowledge and the Flow of Information*. The MIT Press, 1981. (Cited on pages 30, 37, 42, 130, 230, and 231.)

F. Dretske. *Belief, Form, Content and Function*, chapter Misrepresentation, pages 17–36. Oxford University Press, 1986. (Cited on pages 25, 32, 230, and 231.)

F. Dretske. *Explaining Behavior. Reasons in a World of Causes*. The MIT Press, 1988. (Cited on pages 32, 37, 69, 77, 127, and 132.)

F. Dretske. *Naturalizing the Mind*. The MIT Press, 1995. (Cited on pages 37, 129, 150, and 151.)

M. Dummett. *The Seas of Language*. Oxford University Press, 1993. (Cited on page 212.)

D. Earl. Concepts. *The Internet Encyclopedia of Philosophy.*, 2007. (Cited on pages 220, 221, 222, and 227.)

W.G. Eberhard. Aggressive chemical mimicry by a bolas spider. *Science*, 198:1173–1775, 1977. (Cited on page 139.)

H. Eichenbaum; P.Dudchenko; E. Wood; M. Shapiro; H. Tanila. The hippocampus, memory, review and place cells: Is it spatial memory or a memory space? *Neuron*, 23:209–226, 1999. (Cited on pages 266 and 275.)

C. Elder. What versus how in naturally selected representations. *Mind*, 107:349–363, 1998. (Cited on pages 68, 88, and 113.)

C. Eliasmith. *How neurons mean: A neurocomputational theory of representational content*. Unpublished Dissertation, Washington University in St. Louis, 2000. (Cited on pages 36, 39, 40, 163, 164, 165, 167, 174, 176, 177, 179, and 229.)

C. Eliasmith. Moving beyond metaphors: Understanding the mind for what it is. *Journal of Philosophy*, 10:131–159, 2003. (Cited on pages 36 and 211.)

M. Ereshefsky. Species. *Standford Encyclopedia of Philosophy*, 2010. (Cited on page 156.)

D. Hull et al. A general account of selection: Biology, immunology and behavior. *Behavioral and Brain Sciences*, 24(2):511–527, 2001. (Cited on pages 57 and 72.)

E. Kandel et al. *Principles of Neural Science*. McGraw-Hill, 2000. (Cited on pages 164, 166, 167, 174, 266, and 290.)

J. Lettvin et al. What the Frog's Eye Tells the Frog's Brain. *Proceedings of the Institute of Radio Engineers*, 49:1940–1951, 1959. (Cited on page 85.)

G. Evans. *The Varieties of Reference*. Oxford Clarendon Press, 1982. (Cited on pages 154, 197, and 216.)

J. P. Ewert. Neural correlates of key stimulus and releasing mechanism: a case study and two concepts. *Trends in Neuroscience*, 20(8):332–339, 1997. (Cited on pages 185 and 187.)

J. P. Ewert. Motion perception shapes the visual world of amphibians. In F. R. Prete, editor, *Complex Worlds from Simpler Nervous Systems*, pages 117–161. MIT Press, 2004. (Cited on pages 183, 184, 185, 186, and 191.)

J. P. Ewert; H. Buxbaum-Conradi; M. Glagow; A. Roettgen; E. Schuerg Pfeiffer; W. W. Schwippert. Forebrain and midbrain structures involves in prey-catching behavior of toads: Stimulus-response mediating circuits and their modulating loops. *European Journal of Morphology*, 37(2):172–176, 1999. (Cited on pages 183 and 184.)

M. Farah. *The Cognitive Neuroscience of Vision*. Blackwell Publishers, 2000. (Cited on pages 193, 194, 195, and 290.)

H. Field. Mental representation. *Erkenntnis*, 13:9–18, 1978. (Cited on page 127.)

W. Fish. *Philosophy of Perception: a Contemporary Introduction*. Routledge Contemporary Introductions, 2010. (Cited on page 202.)

J. Fodor. *The Language of Thought*. Harvard University Press, Cambridge, 1975. (Cited on pages 72, 216, and 217.)

J. Fodor. Why Paramecia Don't Have Mental Representations. In Uehling French, P., editor, *Midwest Studies in Philosophy*, pages 3–23. University of Minnesota Press, 1986. (Cited on page 216.)

J. Fodor. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press, 1987. (Cited on pages 26, 41, and 271.)

J. Fodor. *A Theory of Content and Other Essays*. The MIT Press, 1990. (Cited on pages 25, 33, 34, 42, 43, 44, 45, 86, 128, 134, 165, 230, and 284.)

J. Fodor. *The Elm and the Expert: Mentalese and Its Semantics*. The MIT Press, 1995. (Cited on pages 212 and 216.)

J. Fodor. *Concepts: Where Cognitive Science Went Wrong*. Oxford Clarendon Press, 1998. (Cited on pages 210, 214, 216, 217, 220, 222, 223, 225, 226, 227, 261, and 294.)

J. Fodor. Language, thought and compositionality. *Cognition*, 16(1): Mind and Language, 2001. (Cited on pages 222, 223, 261, and 262.)

J. Fodor. Having concepts: A brief refutation of the twentieth century. *Mind and Language*, 19(1):29–47, 2004. (Cited on pages 214, 217, 220, and 222.)

J. Fodor. *LOT 2*. Oxford Clarendon Press, 2008. (Cited on pages 26, 44, 210, 214, 215, 216, 217, 224, 226, 249, 252, and 253.)

J. Fodor and E. Lepore. *Holism: A Shopper's Guide*. Blackwell, 1992. (Cited on pages 223 and 294.)

B. Franks and N. Braisby. What is the point? concepts, description, and rigid designation. *Behavioral and Brain Sciences*, 21(1):70–70, 1998. (Cited on page 252.)

G. Frege. *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung ueber den Begriff der Zahl*. Breslau, 1884. (Cited on page 219.)

G. Frege. Ueber sinn un bedeutung. *Zeitschrift fÃŒr Philosophie und philosophische Kritik*, 1892. (Cited on page 212.)

J. Frisby and J. Stone. *Seeing: The Computational Approach to Biological Vision*. The MIT Press, 2010. (Cited on pages 191 and 194.)

C. R. Gallistel. *The Organization of Learning*. MIT press, 1990. (Cited on page 72.)

C. R. Gallistel. *Memory and the computational brain: Why cognitive science will transform neuroscience*. Wiley/Blackwell, 2010. (Cited on page 276.)

D. Garret. Hume's naturalistic theory of representation. *Synthese*, 152 (3):301–319, 2008. (Cited on page 29.)

C. Gauker. Building block dilemmas. *Behavioral and Brain Sciences*, 21 (1):26–27, 1998. (Cited on page 249.)

T. Van Gelder. Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, 14:355–384, 1990. (Cited on page 258.)

S. Gelman. *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford Clarendon Press, Oxford, 2003. (Cited on page 222.)

T. S. Gendler. Why language is not a direct medium. *Behavioral and Brain Science*, 21(1):71–72, 1998. (Cited on page 254.)

D. Gentner. Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj, editor, *Language development: Vol. 2. Language, thought and culture*, pages 301–334. Hillsdale, 1982. (Cited on page 248.)

B. Gertsman. *Epidemology Kept Simple: An introduction to Traditional and Modern Epidemology*. Wiley, New York, 2003. (Cited on page 79.)

M. Ghiselin. A Radical Solution to the Species Problem. *Systematic Zoology*, 23:536–544, 1974. (Cited on page 156.)

A. Glenberg. What memory is for. *Behavioral and Brain Sciences*, 20:1–55, 1997. (Cited on page 277.)

P. Godfrey-Smith. Functions: Consensus without unity. *Pacific Philosophical Quarterly*, 74:196–208, 1993. (Cited on pages 47, 48, 50, and 87.)

P. Godfrey-Smith. A modern history theory of functions. *Nous*, 28(3): 344–362, 1994. (Cited on page 58.)

P. Godfrey-Smith. *Complexity and the Function of Mind in Nature*. Cambridge University Press, 1996. (Cited on pages 51, 73, 97, 101, 103, 104, 126, 129, 134, and 138.)

P. Godfrey-Smith. Mental representation, naturalism and teleosemantics. In MacDonald and D. Papineau, editors, *Teleosemantics*. Oxford University Press, 2006. (Cited on page 73.)

P. Godfrey-Smith. Signals, icons, and beliefs. In D. Ryder, J. Kingsbury, and K. Williford, editors, *Millikan and Her Critics*. Blackwell, 2013. (Cited on pages 69, 70, and 73.)

P.r Godfrey-Smith. *Darwinian Populations and Natural Selection*. Oxford University Press, 2009. (Cited on pages 55, 56, 57, 58, 109, and 138.)

N. Goodman. *Languages of Art*. Hackett Publishing Company, 1976. (Cited on page 29.)

P. Grice. The causal theory of perception. *Proceedings of the Aristotelian Society*, 1:121–153, 1961. (Cited on page 30.)

P. Grice. *Studies in the Way of Words*. Harvard University Press, 1989. (Cited on page 21.)

P. Griffiths. Functional analysis and proper functions. *British Journal for the Philosophy of Science*, 44(3):409–422, 1993. (Cited on page 58.)

P. Griffiths. Squaring the circle: Natural kinds with historical essence. In R. Wilson, editor, *Species: New Interdisciplinary Studies*. MIT Press, 1999. (Cited on page 144.)

J. Hafernik and L. Saul-Gershenz. Beetle larvae cooperate to mimic bees. *Nature*, 6782:35–6, 2000. (Cited on page 142.)

M. Hauser. *The Evolution of Communication*. MIT Press, 1996. (Cited on pages 69 and 74.)

J. Hawthorne. Advice for physicalists. *Philosophical Studies*, 109(1):17–52, 2002. (Cited on page 20.)

D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat striate cortex. *Journal of Physiology*, 148:574–59I, 1959. (Cited on page 175.)

D. Hull. A matter of individuality. *Philosophy of Science*, 45(3):335–360, 1978. (Cited on page 156.)

F. Jackson. *From Metaphysics to Ethics: a Defense of Conceptual Analysis*. Oxford University Press, 1998. (Cited on pages 20 and 219.)

P. Jacob. *What Minds Can Do: Intentionality in a Non-Intentional World*. Cambridge University Press, 1997. (Cited on page 125.)

P. Jacob. Can selection explain content? *The Proceedings of the Twentieth World Congress of Philosophy*, 9:91–102, 2000. (Cited on pages 123 and 125.)

P. Jacob and M. Jeannerod. *Ways of Seeing: The Scope and Limits of Visual Cognition*. Oxford University Press, Oxford, 2003. (Cited on pages 36, 125, 176, 177, 179, 193, 194, and 196.)

R. Jeshion. *New Essays on Singular Thought*. Oxford University Press, Oxford, 2010. (Cited on page 154.)

A. Noe K. O'Reagan. A sensorymotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24:939–1031, 2001. (Cited on page 211.)

E. Kant. *Kritik der reinen Vernunft*. Johann Friedrich Hartknoch, 1787. (Cited on page 219.)

F. C. Keil. *Concepts, Kinds, and Cognitive Development*. MIT Press, 1989. (Cited on page 222.)

C.E-Hani; Joao Keiroz; F. Stjenfelt. Firefly femmes fatales: A case study in the semiotics of deception. *Biosemiotics*, 3(1):33–55, 2010. (Cited on page 142.)

S. Kelly. Demonstrative concepts and experience. *The Philosophical Review*, 397(3):397–420, 2001. (Cited on page 271.)

A. Kenny. Concepts, brains, and behaviour. *Grazer Philosophische Studien*, 21(1):105–113, 2010. (Cited on pages 212 and 213.)

S. Klein; L. Cosmides; J. Tooby; S. Chance. Decisions and the evolution of memory. *Psychological Review*, 109(4):306–329, 2002. (Cited on pages 175, 274, 275, 276, 289, and 290.)

S. Kripke. *Naming and Necessity*. Blackwell, 1980. (Cited on pages 30 and 230.)

S. Kripke. *Wittgenstein on Rules and Private Language*. Harvard University Press, 1982. (Cited on page 112.)

B. Landau. Will the real grandmother please stand up? the psychological reality of dual meaning representations. *Journal of Psycholinguistic Research,*, 11:47–62, 1982. (Cited on page 223.)

S. Laurence and E. Margolis. *Concepts. Core Readings*, chapter Concepts and Cognitive Science. The MIT Press, 1999. (Cited on pages 219, 221, 222, and 225.)

S. Laurence and E. Margolis. Concepts. *Stanford Encyclopedia of Philosophy*, 2011. (Cited on pages 210, 218, 219, 222, 223, 225, 227, 242, and 249.)

S. Laurence and E. Margolis. In defense of nativism. *Philosophical Studies*, Forthcoming. (Cited on pages 214 and 291.)

J. Levine. Swampjoe: Mind or simulation? *Mind and Language*, 11(1): 86–91, 1996. (Cited on pages 154 and 155.)

D. Lewis. *Convention: A Philosophical Study*. John Wiley and Sons, 1969. (Cited on pages 69, 73, 76, and 137.)

D. Lewis. *On the Plurality of Worlds*. Blackwell Publishers, 1986. (Cited on page 237.)

D. Lewis. Reduction of mind. In M. Hahn and B. Ramberg, editors, *A Companion to Philosophy of Mind*. Mit Press, 1994. (Cited on page 22.)

O. Linnebo. Compositionality and frege's context principle. 2008. (Cited on page 262.)

B. Loar. Phenomenal states. *Philosophical Perspectives*, 4:81–108, 1990. (Cited on page 286.)

B. Loar. Phenomenal intentionality as the basis of mental content. In M. Hahn and B. Ramberg, editors, *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. MIT Press, 2003. (Cited on page 296.)

J. Locke. *An Essay Concerning Human Understanding*. The Basset, 1690. (Cited on page 219.)

G. Macdonald and D. Papineau. *Teleosemantics*, chapter Prospects and Problems for Teleosemantics, pages 1–22. Oxford University Press, 2006. (Cited on page 69.)

E. Machery. *Doing Without Concepts*. Oxford University Press, 2009. (Cited on pages 210, 211, 213, 218, 221, 222, 224, and 275.)

B. Mahon and A. Caramazza. Concepts and categories: a cognitive neuropsychological perspective. *Annual Review of Psychology*, 60, pages = 27-51, 2009. (Cited on page 275.)

B. Mahon and A. Caramazza. What drives te organization of object knowledge in the brain? *Trends in Cognitive Science*, 15:97–103, 2011. (Cited on page 289.)

P. Mandik. Varieties of representation in evolved and embodied neural networks. *Biology and Philosophy*, 18:95–130, 2003. (Cited on pages 11, 163, 180, 181, 276, and 278.)

J. M. Mandler. *The Foundations of Mind: Origins of Conceptual Thought*. Oxford University Press, Oxford, 2004. (Cited on pages 217 and 218.)

E. Margolis. How to acquire a concept. *Mind and Language*, 13(3): 347–369, 1998. (Cited on pages 41, 42, 214, 283, and 292.)

E. Margolis and S. Laurence. Learning matters: The role of learning in concept acquisition. *Mind and Language*, 26:507–539, 2011. (Cited on page 283.)

E. Markman. *Categorization and Naming in Children*. Oxford University Press, 1991. (Cited on page 248.)

D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, 1982. (Cited on page 191.)

D.C. Marshall and K. B. R. Hill. Versatile aggressive mimicry of cicadas by an australian preadtory katydid. *PlosOne*, 4(1), 2009. (Cited on page 139.)

M. Martin. The transparency of experience. *Mind and Language*, 4(4): 376–425, 2002. (Cited on pages 202 and 205.)

M. Martinez. *A Naturalistic Account of Content and an Application to Modal Epistemology*. Unpublished dissertation, University of Barcelona, 2010. (Cited on pages 35, 40, 51, 58, 78, 87, 88, 89, 96, 98, 115, 116, 128, 243, 259, 263, and 276.)

M. Martinez. Teleosemantics and productivity. *Philosophical Psychology:*, forthcoming. (Cited on pages 100 and 250.)

M. Matthen. Teleosemantics and the consumer. In G. McDonald and D. Papineau, editors, *Teleosemantics*. Oxford University Press, 2006. (Cited on pages 83 and 166.)

J. Maynard-Smith and D. Harper. *Animal Signals*. Oxford Series in Ecology and Evolution, 2003. (Cited on pages 138, 140, and 143.)

K. B. McDermott, K. K. Szpunar, and S. E. Christ. Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia*, 47:2290–2298, 2009. (Cited on page 290.)

J. McDowell. *Mind and World*. Harvard University Press, 1994. (Cited on pages 165 and 167.)

P. McLaughlin. *What Functions Explain: Functional Explanation and Self-Reproducing Systems*. Cambridge University Press, 2001. (Cited on page 63.)

S. Merilatia. Crypsis through disruptive coloration in an isopod. *Proceedings of the Royal Society of Biological Sciences*, 265(1401):1059–1064, 1998. (Cited on page 143.)

R. Millikan. Reply to antony. In D. Ryder; J. Kingsbury; K. Williford, editor, *Millikan and her critics*. Wiley-Blackwell, 2013. (Cited on page 242.)

R. G. Millikan. *Language, Thought and Other Biological Categories*. The MIT Press, 1984. (Cited on pages 25, 49, 53, 55, 70, 73, 76, 87, 96, 102, 103, 108, 109, 111, 113, 114, 118, 131, 149, 150, 154, 242, 243, 244, 245, 246, and 253.)

R. G. Millikan. In Defense of Proper Functions. *Philosophy of Science*, 56 (2):288–302, 1989. (Cited on pages 48 and 49.)

R. G. Millikan. *White Queen Psychology and Other Essays for Alice*. The MIT Press. Bradford Books, 1993. (Cited on pages 41, 49, 60, 73, 78, 83, 87, 89, 115, 150, 151, and 216.)

R. G. Millikan. Pushmi-Pullyu Representations. *Philosophical Perspectives*, 9:185–200, 1995. (Cited on page 68.)

R. G. Millikan. On swampkinds. *Mind and Language*, 11(1):103–17, 1996. (Cited on pages 150 and 155.)

R. G. Millikan. 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, 1998. (Cited on pages 115, 116, 117, 223, 244, 246, 247, 248, 254, 284, and 292.)

R. G. Millikan. Wings, spoons, pills, and quills: A pluralist theory of function. *Journal of Philosophy*, 96(4):191–206, 1999. (Cited on page 21.)

R. G. Millikan. *On Clear and Confused Ideas*. Cambridge University Press, 2000. (Cited on pages 144, 198, 212, 213, 216, 224, 226, 237, 242, 243, 244, 245, 246, 247, 250, 254, 259, 264, 278, 284, and 292.)

R. G. Millikan. *Functions: New Essays in the Philosophy of Psychology and Biology*, chapter Biofunctions: Two Paradigms, pages 113–143. Oxford University Press, 2002. (Cited on pages 49, 70, 113, and 116.)

R. G. Millikan. *Varieties of Meaning*. London: MIT Press, 2004. (Cited on pages 32, 42, 47, 68, 72, 76, 77, 84, 88, 102, 131, 133, 137, 140, 168, 186, 193, 254, and 260.)

R. G. Millikan. *Language: A Biological Model*. The MIT Press, 2005. (Cited on pages 21, 55, 76, 137, 243, and 247.)

R. G. Millikan. An Input Condition for Teleosemantics? A reply to Shea (and Godfrey-Smith). *Philosophy and Phenomenological Research*, 75(2), 2007. (Cited on pages 127 and 132.)

M. Mossio and A. Moreno. Organisational closure in biological organisms. *History and Philosophy of the Life Sciences*, 32(2-3):269–288, 2010. (Cited on page 124.)

M. Mossio, C. Saborido, and A. Moreno. An organizational account of biological functions. *British Journal for the Philosophy of Science*, 60(4): 813–841, 2009a. (Cited on pages 51, 59, 62, 63, and 64.)

M. Mossio, C. Saborido, and A. Moreno. El concepto de funcion biologica desde un enfoque organizacional. *Actas del VI Congreso de la Sociedad de Logica, Metodologia y Filosofia de la Ciencia en Espana*, pages 553–558, 2009b. (Cited on pages 59 and 63.)

M. Mossio, C. Saborido, and A.o Moreno. Biological organization and cross-generation functions. *British Journal for the Philosophy of Science*, 62(3):583–606, 2011. (Cited on pages 65 and 66.)

G. Murphy. *The Big Book if Concepts*. MIT Press, 2002. (Cited on pages 210 and 222.)

K. Neander. Functions as Selected Effects: The Conceptual Analyst's Defence. *Philosophy of Science*, 58:168–184, 1991. (Cited on pages 48, 53, and 156.)

K. Neander. Misrepresenting & Malfunctioning. *Philosophical Studies*, 79:109–141, 1995. (Cited on pages 73, 87, 88, 121, 125, and 150.)

K. Neander. Swampman meets swampcow. *Mind and Language*, 11(1): 118–29, 1996. (Cited on page 156.)

K. Neander. Types of traits: The importance of functional homologues. In R. Cummins A. Ariew and M.Perlman, editors, *Functions: New Readings in the Philosophy of Psychology and Biology*. Oxford University Press, 2002. (Cited on page 156.)

K. Neander. Content for cognitive science. In G. MacDonald and D. Papineau, editors, *Teleosemantics*. Oxford University Press, 2006. (Cited on pages 34, 85, 87, 88, 125, 165, 179, and 183.)

K. Neander. Teleological theories of mental content. *Standford Encyclopedia of Philosophy*, 2012. (Cited on pages 32 and 240.)

K. Neander. Toward an informational teleosemantics. In D. Ryder; J.Kingsbury; K. Williford, editor, *Millikan and her critics*. Wiley-Blackwell, 2013. (Cited on pages 129, 130, and 131.)

D. Osherson and E. Smith. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9(1):25–58, 1981. (Cited on pages 221, 222, and 223.)

L. Darden P. Machamer and C. F. Craver. Thinking about mechanisms. *Philosophy of Science*, 67:1–25, 2000. (Cited on page 73.)

D. Papineau. Representation and explanation. *Philosophy of Science*, 5 (4):550–572, 1984. (Cited on pages 69, 149, and 150.)

D. Papineau. *Reality and Representation*. Basil Blackwell, 1987. (Cited on pages 23, 25, and 69.)

D. Papineau. *Philosophical Naturalism*. Basil Blackwell, 1993. (Cited on pages 19, 47, 69, 73, 87, 131, 151, and 241.)

D. Papineau. Teleosemantics and Indeterminacy. *Australasian Journal of Philosophy*, 76(1):1–14, 1998. (Cited on pages 124, 150, 165, 238, and 239.)

D. Papineau. The Status of Teleosemantics, or How to Stop Worrying About Swampman. *Australasian Journal of Philosophy*, 79(2):279–89, 2001. (Cited on pages 150, 151, and 152.)

D. Papineau. Is representation rife? *Ratio*, 16(2):107–123, 2003. (Cited on pages 68, 77, 78, 113, 284, and 292.)

D. Papineau. *The Oxford Handbook of Philosophy of Language*, chapter Naturalist Theories of Meaning, pages 175–188. Oxford University Press, 2006a. (Cited on page 257.)

D. Papineau. Phenomenal and perceptual concepts. In T.Alter and S. Walter, editors, *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press, 2006b. (Cited on pages 285 and 286.)

D. Papineau. Naturalism. *Stanford Encyclopedia of Philosophy*, 2007. (Cited on page 19.)

David Papineau. Doubtful intuitions. *Mind and Language*, 11(1):130–2, 1996. (Cited on page 154.)

A. Pautz. What are the contents of experiences? *Philosophical Quarterly*, 59(236):483–507, 2008. (Cited on page 202.)

Christopher Peacocke. *A Study of Concepts*. The MIT Press, 1992. (Cited on pages 202, 211, 212, 213, and 217.)

D. Pineda. *La Mente Humana. Introduccion a la Filosofia de la Psicologia*. Catedra, 2012. (Cited on pages 28 and 29.)

D. Pitcher; L.Charles; J. T. Delvin; V. Walsh; B. Duchaine. Triple dissociation of faces, bodies and objects in extrastriate cortex. *Current Biology*, 19:319–324, 2009. (Cited on page 290.)

D. Pitt. The phenomenology of cognition or what is it like to think that p?r. *Philosophy and Phenomenological Research*, 69:1–36, 2004. (Cited on page 271.)

B. Preston. Why is a wing like a spoon? a pluralist theory of function. *The Journal of Philosophy*, 95(5):215–254, 1998. (Cited on pages 113 and 115.)

C. Price. Determinate Functions. *Noûs*, 32(1):54–75, 1998. (Cited on page 87.)

C. Price. *Functions in Mind*. Oxford University Press, Oxford, 2001. (Cited on pages 87 and 131.)

J. Prinz. The duality of content. *Philosophical Studies*, 100(1):1–34, 2000. (Cited on pages 87, 230, and 236.)

J. Prinz. *Furnishing the Mind: Concepts and their perceptual basis*. MIT Press, 2002. (Cited on pages 27, 28, 29, 30, 129, 210, 217, 218, 221, 222, 224, 225, 226, 227, 230, 231, 232, 233, 237, 275, 284, and 289.)

J. Prinz. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford University Press, 2004. (Cited on page 230.)

J. Prinz. Beyond appearances : The content of sensation and perception. In T. S. Gendler and J. Hawthorne, editors, *Perceptual Experience*. Oxford University Press, 2006. (Cited on page 230.)

J. Prinz. Regaining composure: A defense of prototype compositionality. In W. Hinzen M. Werning and E. Machery, editors, *The Oxford Handbook of Compositionality*. Harvard University Press, 2008. (Cited on page 222.)

Z. Pylyshyn. Is vision continuous with cognition? the case for cognitive impenetrability of visual perception? *Behavioral and Brain Sciences*, 22 (3):341–365, 1999. (Cited on pages 194 and 197.)

Z. Pylyshyn. *Seeing and Visualizing: It's Not What You Think*. The MIT Press, 2003. (Cited on pages 29, 167, 169, 194, 197, and 228.)

Z. Pylyshyn. Some puzzling findings in multiple object tracking: I. tracking without keeping track of object identities. *Visual Cognition*, 11(7):801–822, 2004. (Cited on pages 194 and 198.)

Z. Pylyshyn. *Things and Places: How the Mind Connects with the World*. The MIT Press, 2007. (Cited on pages 29, 194, and 197.)

W. V. O. Quine. Two dogmas of empiricism. In *From a Logical Point of View*. Harvard University Press, 1953. (Cited on page 220.)

D Radner. Mind and function in animal communication. *Erkenntnis*, 51: 129–144, 1999. (Cited on page 140.)

A. Raftopoulos. *Cognition and Perception: How do Psychology and the Neural Science inform Philosophy*. The MIT Press, 2009a. (Cited on pages 30, 167, 193, 194, 195, 196, 197, and 198.)

A. Raftopoulos. Reference, perception, and attention. *Philosophical Studies*, 144:339–360, 2009b. (Cited on page 193.)

F. Recanati. Perceptual concepts: in defence of the indexical model. *Synthese*, Forthcoming. (Cited on pages 198 and 283.)

M. Rescorla. Millikan on honeybee navigation and communication. In D. Ryder; J. Kingsbury; K. Williford, editor, *Millikan and her Critics*. Blackwell, 2013. (Cited on pages 131 and 169.)

G. Rey. Concepts and conceptions: a rely to smith, medin and rips. *Cognition*, 15:237–62, 1983. (Cited on page 222.)

G. Rey. Concepts and stereotypes. *Cognition*, 19:297–303, 1985. (Cited on pages 210, 213, and 220.)

P. Robbins. The myth of reverse compositionality. *Philosophical Studies*, 125(2):251–275, 2005. (Cited on pages 222 and 262.)

E. Rosch. Principles of categorization. In E. Rosch and B. B. Lloyd, editors, *Cognition and categorization*. Hillsdale, Erlbaum, 1978. (Cited on pages 220 and 221.)

E. Rosch and C. B. Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975. (Cited on page 221.)

R. D. Rupert. Causal theories of mental content. *Philosophy Compass*, 3: 353–380, 2008. (Cited on pages 33, 34, 39, 40, and 231.)

G. Ruxton; T. Sherratt; M. Speed. *Avoiding Attack. The evolutionary ecology of crypsis, warning signals, and mimicry*. Oxford University Press, 2004. (Cited on pages 139, 140, 142, and 143.)

D. Ryder. On thinking of kinds: a neuroscientific perspective. In G. MacDonald and D. Papineau, editors, *Teleosemantics*. Oxford University Press, 2006. (Cited on pages 33 and 165.)

D. Ryder. Problems of representation ii: Naturalizing content. In F. Garzon and J. Symons, editors, *Teleosemantics*. Routledge Companion to the Philosophy of Psychology, 2009. (Cited on pages 163, 165, 167, and 210.)

D. Ryder. Too close for comfort? psychosemantics and the distal. *Personal website*, Unpublished. (Cited on page 297.)

L. Gleitman (1983) S. Armstrong, H. Gleitman. What concepts might not be. *Cognition*, 13:3:263–308, 1983. (Cited on pages 219 and 222.)

R. Sainsbury and M. Tye. *Seven Puzzles of Thought and How to Solve Them: An Originalist Theory of Concepts*. OUP, 2012. (Cited on page 212.)

K. O. Salomon, D. L. Medin, and E. B: Lynch. Concepts do more than categorize. *Trends in Cognitive Science*, 3(3):99–105, 1999. (Cited on page 213.)

I. Sazima; L. Nobre-Carvalho; F. Pereira Mendonca; J. Zuanon. Fallen leaves on the water-bed: diurnal camouflage of three night active fish species in an amazonian streamlet. *Neotropical Ichthyology*, 41(1), 2006. (Cited on page 143.)

D. Schachter. *The Seven Sins of Memory*. Houghton Mifflin Company, 2001. (Cited on pages 275 and 276.)

S. Schellenberg. The particularity and phenomenology of perceptual experience. *Philosophical Studies*, 149(1):19–48, 2010. (Cited on pages 76, 137, 155, 202, 204, 205, and 206.)

S. Schellenberg. Perceptual content defended. *NoÃ»s*, 45(4):714–750, 2011. (Cited on pages 202 and 204.)

S. Schneider. *The Language of Thought A New Philosophical Direction*. MIT Press, 2011. (Cited on pages 225 and 227.)

B. Scholl. Objects and attention: the state of the art. *Cognition*, 80:1–46, 2001. (Cited on page 194.)

J. Schroeder. Explanatory force, antidescriptionism, and the common structure of substance concepts. *Behavioral and Brain Sciences*, 21(1): 84–85, 1998. (Cited on page 224.)

W. Searcy and S. Nowicki. *The Evolution of Animal Communication*. Princeton University Press, 2005. (Cited on page 137.)

J. Searle. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press, 1983. (Cited on page 21.)

M. A. Sebastian. *Self-Involving Representationalism: a naturalistic theory of phenomenal consciousness*. Unpublished Dissertation, 2011. (Cited on page 151.)

C. Semenza. The neuropsychology of proper names. *Mind and Language*, 24(4):347–369, 2009. (Cited on page 290.)

C. Semenza and M. Zettin. Generating proper names: A case of selective inability. *Cognitive Neuropsychology*, 5(6):711–721, 1988. (Cited on page 290.)

R. M. Seyfarth; D. Cheney; T. Bergman; J. Fischer; K. ZuberbÃŒhler; K. Hammerschmidt. The central importance of information in studies of animal communication. *Animal Behaviour*, 80:3–8, 2010. (Cited on page 138.)

N. Shea. Consumers Need Information: Supplementing Teleosemantics with an Input Condition. *Philosophy and Phenomenological Research*, 75 (2):404–435, 2007. (Cited on pages 126, 127, 128, 131, 135, and 136.)

N. Shea. Millikan's isomorphism requirement. In D. Ryder; J. Kingsbury; K. Williford, editor, *Millikan and Her Critics*. Blackwell, 2013. (Cited on page 131.)

Nicholas Shea. Millikan's isomorphism requirement. In *Millikan and Her Critics*. forthcoming. (Cited on page 110.)

D. Sherry and D. Schachter. The evolution of multiple memory systems. *Psychological Review*, 94(4):439–454, 1987. (Cited on pages 176, 276, and 289.)

S. Shettleworth. *Cognition, Evolution, and Behavior*. Oxford University Press, Oxford, 2010. (Cited on pages 74, 97, 276, and 278.)

P. Shulte. How frogs see the world: Putting millikan's teleosemantics to the test. *Philosophia*, 40:483–496, 2012. (Cited on page 125.)

S. Siegel. *The Contents of Visual Experience*. Oxford University Press, Oxford, 2011. (Cited on pages 202 and 204.)

N. Sinclair. Methaetics, teleosemantics and the content of moral judgement. *Biology and Philosophy*, forthcoming. (Cited on pages 113 and 116.)

P. Singer. *Animal Liberation: A New Ethics for our Treatment of Animals*. New York Review/Random House, 1975. (Cited on page 152.)

A. Sirovic; J. Hildebrand; S.Wiggins; M.McDonald; S.Moore; D. Thiele. Seasonality of blue andfin whale calls andthe influence of sea ice in the western antarctic peninsula. *Deep-Sea Research*, 51:2327–2344, 2004. (Cited on page 129.)

B. Skyrms. *Evolution of the Social Contract*. Cambridge University Press, 1996. (Cited on pages 73, 101, 137, and 138.)

B. Skyrms. *Signals: Evolution, Learning, and Information*. Oxford University Press, Oxford, 2010. (Cited on pages 42, 69, 73, 76, 101, 138, and 180.)

E. E. Smith and D. L. Medin. *Categories and concepts*. Cambridge University Press, 1981. (Cited on pages 220 and 251.)

E. Sober. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. University of Chicago Press, 1984. (Cited on pages 57, 133, and 144.)

E. Sober. The two faces of fitness. In R. Singh; D. Paul; C. Krimbas; J. Beatty, editor, *Thinking about Evolution: Historical,Philosophical, and Political Perspectives*, pages 309–321. Cambridge University Press, 2002. (Cited on page 133.)

R. Sorabji. Function. *Philosophical Quarterly*, 14(57):289–302, 1964. (Cited on page 60.)

M. Soteriou. The particularity of visual perception. *European Journal of Philosophy*, 8(2):173–189, 2000. (Cited on page 205.)

M. Staaden; H. Roemer; V. Couldridge. A novel approach to hearing: The acoustic world of pneumoid grasshoppers. In M. R. Prete, editor, *Complex Worlds from Simple Nervous Systems:*. MIT Press, 2004. (Cited on page 188.)

R. Stalnaker. *Inquiry*. MIT Press, 1984. (Cited on page 41.)

D. Stampe. Towards a Causal Theory of Linguistic Representation. *Midwest Studies in Philosophy*, 2:42–63, 1977. (Cited on pages 30 and 31.)

U. Stegmann. John maynard's smith notion of animal signals. *Biology and Philosophy*, 20:1011–1025, 2005. (Cited on pages 68 and 139.)

U. Stegmann. A consumer based teleosemantics for animal signals. *Philosophy of Science*, 76(5), 2009. (Cited on pages 73, 139, 141, 144, 147, and 148.)

K. Sterelny. *The Representational Theory of Mind: An Introduction*. Oxford University Press, 1990. (Cited on pages 87, 165, 210, and 284.)

K. Sterelny. Basic Minds. *Philosophical Perspectives*, 9, AI, Connectionism and Philosophical Psychology:251–270, 1995. (Cited on pages 26, 40, 139, 141, and 172.)

K. Sterelny. *Thought in a Hostile World*. Blackwell Publishing, 2003. (Cited on pages 166, 169, 201, 259, and 279.)

K. Sterelny and P. Griffiths. *Sex and Death: An Introduction to Philosophy of Biology*. University of Chicago Press, 1999. (Cited on pages 56, 57, 140, 141, and 156.)

R. Sternberg. *Cognitive Psychology*. Wadsworth, 2009. (Cited on pages 25, 29, 164, 275, and 289.)

S. Stich. *From Folk Pschology to Cognitive Science: The Case Against Belief*. MIT Press, 1983. (Cited on page 169.)

D. Stoljar. Physicalism. *Stanford Encyclopedia of Philosophy*, 2009. (Cited on page 20.)

D. Stoljar. *Physicalism*. Routledge, 2010. (Cited on pages 20 and 22.)

D. Stuart-Fox and A. Moussalli. Selection for social signalling drives the evolution of chameleon colour change. *PLoS Biology*, 6(1):e25, 2008. (Cited on page 76.)

A. M. Sullivan; J. C. Maerz and D. M. Madisoni. Anti-predator response of red-backed salamanders (plethodon cinereus) to chemical cues from garter snakes (thamnophis sirtalis): Laboratory and field experiments. *Behavioral Ecology and Sociobiology*, 51(3):227–233, 2001. (Cited on page 82.)

Z. G. Szabo. Compositionality. *The Stanford Encyclopedia of Philosophy (Winter 2008 edition)*, 2007. URL `http://plato.stanford.edu/archives/win2008/entries/compositionality/`. (Cited on pages 261 and 262.)

K. Taylor. On singularity. In R. Jeshion, editor, *New Essays on Singular Thought*. Oxford University Press, Oxford, 2010. (Cited on page 265.)

N. Tinbergen. *The Herring Gull's World*. Basic Books, New York, 1960. (Cited on pages 40 and 185.)

P. Toribio. State versus content: The unfair trial of perceptual nonconceptualism. *Erkenntnis*, 69(3):351–361, 2008. (Cited on page 204.)

M. Tye. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. The MIT Press. Bradford Books, 1995. (Cited on page 202.)

M. Tye. *Consciousness, Color and Content*. The MIT Press. Bradford Books, 2000. (Cited on page 202.)

M. Tye. *Consciousness Revisited: Materialism Without Phenomenal Concepts*. The MIT Press, 2009a. (Cited on pages 20, 202, 204, 286, and 287.)

M. Tye. The admissible contents of visual experience. *Philosophical Quarterly*, 59(236):541–562, 2009b. (Cited on page 204.)

P. Kroes U. Krohs. Philosophical perspectives on organismic and artifactual functions. In P. Kroes U. Krohs, editor, *Functions in Biological and Artificial Worlds: Comparative Philosophical Perspectives*. MIT Press, 2009. (Cited on page 51.)

E. Warrington and T. Shallice. Category specific semantic impairments. *Brain*, 107:829–854, 1984. (Cited on page 290.)

W. Watkins; P. Tyack; K. E. Moore; J. Bird. The 20-hz signal of finback whales (balaenoptera physalus). *The Journal of the Acoustical Society of America*, 82:1901–1912, 1987. (Cited on page 129.)

J. Weinberg. Making sense of empiricism? *Metascience*, 12:279–303, 2003. (Cited on page 224.)

D. A. Weiskopf. The origins of concepts. *Philosophical Psychology*, 21(2): 251–268, 2008. (Cited on page 226.)

D. A. Weiskopf. Atomism, pluralism, and conceptual content. *Philosophy and Phenomenological Research*, 79(1):131–163, 2009. (Cited on pages 215, 217, 218, 224, and 225.)

R. Wilson. *Species: New Interdisciplinary Essays*. MIT Press, 1999. (Cited on pages 144 and 156.)

L. Wittgenstein. *Philosophical Investigations*. Prentice Hall, 1953. (Cited on page 222.)

A. Wouters. The function debate in philosophy. *Acta Biotheoretica*, 53(2): 123–151, 2005. (Cited on pages 50, 51, and 59.)

L. Wright. Functions. *Philosophical Review*, 82(2):139–168, 1973. (Cited on pages 51, 54, and 60.)

A. Zahavi and A. Zahavi. *The Handicap Principle*. Oxford University Press, 1997. (Cited on page 138.)

R. Zahn; J. Moll; F. Krueger; E. Huey; G. Garrido; J. Grafman. Mate selection: A selection for a handicap. *Journal of Theoretical Biology*, 53: 205–14, 1975. (Cited on page 138.)

R. Zahn; J. Moll; F. Krueger; E. Huey; G. Garrido; J. Grafman. Social concepts are represented in the superior anterior temporal cortex. *Preceedings of the National Academy of Sciences*, 104(15):6430–6435, 2007. (Cited on pages 289 and 290.)

E. Zalta. Fregean senses, modes of presentation, and concepts. *Philosophical Perspectives*, 15(15):335–359., 2001. (Cited on page 211.)