2011

In Favor of Teleosemantics: A Millikanian Treatment of the Intentional Content of Mental Representation

Pierre Faye
Graduate Center, City University of New York

IN FAVOR OF TELEOSEMANTICS:

A Millikanian Treatment of the Intentional Content of Mental Representation

by

PIERRE FAYE

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of
the requirements for the degree of Doctor of Philosophy,
The City University of New York

2011

This manuscript has been read and accepted for the
Graduate Faculty in Philosophy in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Date	Chair of Examining Committee John Greenwood
Date	Executive Officer Iakovos Vasiliou

Michael Devitt (Advisor)

Ruth Garrett Millikan

John Greenwood

Richard L. Mendelsohn

Steven Ross
Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

IN FAVOR OF TELEOSEMANTICS:

A Millikanian Treatment of the Intentional Content of Mental Representation

by

Pierre Faye

Advisor: Professor Michael Devitt

Theoretical attempts to naturalize mental contents, that is, to explain how wholly physical organisms manage to represent the external world to themselves, are mostly conducted in accordance with causal-informational and/or functionalist approaches based on nomic physical correlations. In 1984, Ruth Millikan and David Papineau simultaneously, though independently, injected new life into the naturalist program by introducing a divergent approach known today as "teleosemantics."

In first approximation, teleosemantics purports to naturalize mental content by substituting for the former concept of nomic correlations found in causal and/or functionalist models, the biological concept of etiological functions resulting from natural selection. Since its introduction, teleosemantics has been an object of constant misunderstanding and resolute opposition. The goal of this dissertation is to demonstrate that, when properly conceived, teleosemantics is indeed a coherent project capable of responding to the central objections raised against it. Offering a defense of teleosemantics is of critical value to the general program of naturalization of mental content because the spirit of the teleosemantic approach resonates best with the deepest

philosophical tenets of the naturalist enterprise. I want to argue that only a teleological perspective, that is an analysis of etiological functions grounded in the actual history of selected beneficial mechanisms for generating mental representations, is able to explain the real nature of intentional content. Millikan's models of teleosemantics will function as my main frame of reference: her model represents the best contemporary program of intentional realism developed in strictly naturalist terms. This dissertation develops into four chapters. Chapters one and two present a criticism of causal-functionalist models and an analysis of their inability to overcome the challenge of misrepresentation, giving reasons to look for an alternative perspective. Chapter three introduces teleosemantics as a potential candidate for such an alternative model, focusing on Ruth Millikan's perspective, with the ambition to alleviate the many misunderstandings and confusions generally attached to this view. Chapter four addresses the apparently powerful objections against the historical dimension of teleological functions and the controversial role this historical dimension is supposed to play in fixing the intentional content of mental representations in teleosemantics.

In memory of my father.

To Anne-Marie, Caroline and Mylène.

Acknowledgments

I wish to express my sincere gratitude to my advisor, Prof. Michael Devitt, not only for his constant support both on intellectual matters and practical issues, but also for his confidence in my ability to complete this ambitious project.

I would like to thank Prof. Ruth Millikan, whose impressive philosophical work plays a central part in this dissertation. With incredible commitment and generosity, Prof. Millikan helped me understand her teleological model thanks to her careful comments and enlightening explanations.

I have been extremely fortunate to benefit from Prof. Michael Devitt and Prof. Ruth Millikan's guidance during all these years.

One of the first courses I took at the Graduate Center was Prof. Alberto L. Cordero's Philosophy of Science class. Since then, Prof. Cordero has become a mentor and a friend. He introduced me to the subtleties of quantum mechanics and relativity. I am thankful for the many wonderful evenings of passionate philosophical discussions and exquisite meals.

Finally, I want to thank my friend Jeffrey Stephenson. Jeffrey has always been there for me when I needed to improve the quality of my writing, thanks to his great mastery of English language. I am also grateful for his helpful philosophical comments and valuable insights.

IN FAVOR OF TELEOSEMANTICS

TABLE OF CONTENTS

INTRODUCTION

This dissertation is a research in philosophy of mind conducted within the framework of naturalism. It aims at providing a naturalist account for the intentional content of mental representations. Naturalist philosophers want to explain mental activity as the result of entirely natural processes, in particular, although not necessary exclusively, causal processes in the brain. They reject Spiritualism, Vitalism and other forms of substance-dualism. These are also thinkers whose rejection of the distinction between metaphysics and natural philosophy leads to embrace the best knowledge of the time, the one produced by the sciences, as a guiding line for their research.

Theoretical attempts to naturalize mental contents, that is, to explain how wholly physical organisms manage to represent the external world to themselves, are mostly conducted in accordance with causal-informational and/or functionalist approaches based on nomic physical correlations. In 1984, Ruth Millikan and David Papineau simultaneously, though independently, injected new life into the naturalist program by introducing a divergent approach known today as "teleosemantics." Many other philosophers, notably Karen Neander and Peter Godfrey-Smith, have contributed to this original perspective, providing substantial modifications and improvements. Steven Wagner suggested the neologism "teleosemantics" to Millikan (1993) as a short-cut for "a teleological account of what determines semantic contents of inner representations" (p. 123).

In first approximation, teleosemantics purports to naturalize mental content by substituting for the former concept of nomic correlations found in causal and/or functionalist models, the biological concept of etiological functions resulting from natural selection. This new approach quickly gained ground. Thus, in the face of the many challenges encountered by his indicator semantics approach, Dretske (1988) decided to evolve his purely informational model into a hybrid one which encompassed teleosemantic elements. At some point, Jerry Fodor himself, today a fierce opponent of this way of analyzing intentional content was engaged in teleosemantic thinking.[1]

Since its introduction, teleosemantics has also been an object of constant misunderstanding and resolute opposition. For the last twenty years, proponents have been active, refining their models in response to the challenges of the successive waves of increasingly elaborate objections. As a result, the terms of the naturalist debate have substantively changed. Nonetheless, the solutions offered by teleosemanticists in light of the objections have generally left skeptics unmoved. Worse, reflecting on the general attempt to naturalize semantic properties of mental representation, Godfrey-Smith (2006) recently noted that "this whole program seems to have lost momentum, at least for now," and that "the teleosemantic program is not insulated from the general turn away from optimism" (pp. 42-43). He then concluded that a new look at the basic notions involved in theoretical model of mental representation was needed for the program to regain its philosophical force.

[1] Fodor considers teleology with an open mind in *Semantics, Wisconsin Style* (1984, pp. 231-250), and in *Fodor's Guide to Mental Representation* (Spring 1985, pp. 55-97). Fodor's version of teleosemantics, which he rejected later on, can be found in *Psychosemantics or: Where Do Truth Conditions Come From?* (1999).

The goal of this dissertation is to contribute to such an effort by harnessing the explanatory power of Darwinism in relation to teleosemantics.[2] I aim to demonstrate that, when properly conceived, teleosemantics is indeed a coherent project capable of responding to the central objections raised against it. Offering a defense of teleosemantics is of critical value to the general program of naturalization of mental content because the spirit of the teleosemantic approach resonates best with the deepest philosophical tenets of the naturalist enterprise.

I want to argue that only a teleological perspective, that is an analysis of etiological functions grounded in the actual history of selected beneficial mechanisms for generating mental representations, is able to explain the real nature of intentional content. Millikan's models of teleosemantics will function as my main frame of reference: her model represents the best contemporary program of intentional realism developed in strictly naturalist terms. However, many other theorists have offered different versions of teleosemantics that deserved to be considered and thus will be explored as well.

A considerable discrepancy exists between the radically new perspective offered by the naturalist program and the old-fashioned way of arguing for and against it. In

[2] What is meant here by "Darwinism" is clearly specified by Ernst Mayr's following remarks: "Darwin's views on evolution are often referred to as The Darwinian Theory. Actually they consist of a number of different theories that are best understood when clearly distinguished from each other.… Two of these five theories, evolution as such and the theory of common descent, were widely accepted by biologists within a few years of the publication of the *Origin*. This represented *the first Darwinian revolution*. The acceptance of man as a primate in the animal kingdom was a particularly revolutionary step. Three other theories, gradualism, speciation, and natural selection, were strongly resisted and were not generally accepted until the evolutionary synthesis. This was *the second Darwinian Revolution.* The Darwinism proposed by Weismann and Wallace, in which an inheritance of acquired characters is rejected, was named *Neodarwinism* by George Jones Romanes. The Darwinism accepted since the evolutionary synthesis is best simply called *Darwinism*, because in most crucial aspects it agrees with the original Darwinism of 1859, while the belief in an inheritance of acquired characters is by now totally obsolete." (Mayr, 2001, pp. 86-87).

particular, most of the objections raised against teleosemantics, I will argue, are ultimately unsound because they spring from the unreflective endorsement of ill-formed traditional notions such as the ones of “meaning”, “representation”, or “propositional attitude”. This is why the dialogue with opponents is so difficult. When faced with objections infused with preconceptions about intentional content the naturalist thinker is pressured to contrive his theory to make it coincide with theoretical expectations and semantic intuitions set by pre-evolutionist philosophies. By offering a new reading of the teleosemantic approach I hope to provide naturalist philosophers with compelling arguments to build a convincing case for shifting to an evolutionary perspective about mental content, one in resonance with the general spirit of teleosemantics.

Thus, philosophers must begin to understand the central role that Darwinian evolution should play in their thinking. Evolution is, of course, primarily a fact about the history of life on earth. Life has evolved from primitive cells into complex organisms, and Darwinism is the theory specifying the mechanisms responsible for this evolutionary process. The main achievement of Darwinism is its demonstration of how extremely complex organisms fine-tuned to their environment and capable of intelligent behaviors have been the emergent result of blind mechanisms operating over geological time through gradual modifications. In light of such considerations, one would expect evolutionary biology to have had a tremendous impact on the project of naturalizing the mind. Yet it is a disconcerting fact that most of the naturalist models of intentional content have remained aloof of Darwinian considerations. The functionalist framework that supports the vast majority of the predominant models in the literature, for example, yields a kind of design that renders the evolutionary history of representational mechanisms—as well as the historical purpose of the representations for which such mechanisms are responsible—largely irrelevant to the explanation of intentional contents.

Most philosophers have not yet acknowledged the power of evolutionary thinking and the intellectual revolution that it commands in other fields of inquiry besides biology. Providing theoretical models of intentionality merely compatible with or even supported by empirical imports from evolutionary biology will not suffice. To entertain any hope of succeeding in naturalizing intentional content, thinkers need to drastically revise their modes of reasoning. They must embrace evolutionary thinking as a philosophical perspective and let it entirely reshape their theoretical models.

This dissertation develops into four chapters. Chapters one and two present a criticism of causal-functionalist models and an analysis of their inability to overcome the challenge of misrepresentation, giving reasons to look for an alternative perspective. Chapter three introduces teleosemantics as a potential candidate for such an alternative model, focusing on Ruth Millikan's perspective, with the ambition to alleviate the many misunderstandings and confusions generally attached to this view. Chapter four addresses the apparently powerful objections against the historical dimension of teleological functions and the controversial role this historical dimension is supposed to play in fixing the intentional content of mental representations in teleosemantics. In the eyes of its opponents, such objections still constitute today the fundamental reason for rejecting a teleological treatment of mental representations.

Before moving to the first chapter however, one particular clarification is needed due to the fact that no consideration will be given to the problem of consciousness in this work. The following section presents some arguments supporting the idea that consciousness and intentionality are two distinct issues which do not suffer, and in fact could well benefit, from being treated separately.

Treating Intentionality Aside From Consciousness: Some Justifications

The mind is typically regarded as capable of consciousness and intentionality. This work touches upon the latter only. The question to decide whether or not intentionality should or could be subjected to philosophical inquiry for its own sake and quite aside from any considerations about consciousness receives conflicting answers in the literature. It seems however that in recent years more and more thinkers came to regard a separate treatment of the two as uncontroversial.

At least, it is worth noticing that the way most contemporary models of mental representation are set up makes an independent treatment of intentionality both technically possible and theoretically justified, whether or not the authors of such models reflectively acknowledge this to be the case. Any functionalist model for example has such consequences. Functionalists are usually well aware of this fact and sometimes make it central to their position. Hence, some philosophers, while providing a functionalist account of intentionality, have strong reservations about applying the same strategy to consciousness, for which, they sometimes argue, a strictly physicalist approach is to be preferred.[3]

Most of the experts support the idea of studying intentionality independently of consciousness, even when they offer different approaches to the issue; some of them go a step further. David Rosenthal (1999), for example, who has dedicated a large part of his research to consciousness and the defense of the Higher Order Though (HOT hereafter)

[3] Ned Block (1978; 1999), for example, defends such a distinction. Block's view can be regarded as sound only under the assumption that intentionality and consciousness can indeed be studied apart from one another. This assumption is also shared by non-functionalist naturalist as well. Millikan, for example, often jokes that she has kept the question of consciousness for another life, tackling the question on intentionality in this one being quite enough a task.

perspective not only supports the idea that consciousness and intentionality should be approached as separate questions, but argues that if consciousness “were intrinsic to sensory or intentional character, no theoretical understanding of what is to be a conscious state would be possible at all” (pp. 729-753). Rosenthal reasoning can be recapped as follows: one can expect the explanation of the fact that a certain mental state S is a conscious state to appeal to other mental states. Under the assumption that every mental state is conscious, such an explanation will turn out to be circular or at least not informative, referring to consciousness to explain consciousness.

Supporters of qualia will feel uneasy with the idea of explaining mental states in strictly non-mental terms, as naturalism requires, for reasons having to do with the supposed explanatory gap between the mental and the physical (Jackson, 1986; Levine, 1983; 1993). It is one of the great merits of the HOT theory supported by Rosenthal to ease the transition from conscious mental states to their ultimate non mental explanation. Hence the fact that S is conscious is explained by the existence of S’, a non conscious yet intentional state about S. S’ being itself eventually explainable in non mental terms. Here again the explanation requires that intentionality and consciousness be treated independently of one another.

A number of scientific considerations support this contemporary trend in philosophy of mind as well. Thanks to the advances of cognitive sciences and experimental psychology in recent years, an impressive set of experimental data has been piling up, comforting philosophers in adopting this division of labor in the study of intentionality and consciousness. Experiments conducted on brain damaged patients with visual impairment provide especially valuable insights.

There is a large variety of well-documented neuropsychological syndromes in which an act of vision is successfully accomplished without being accompanied by any awareness of seeing. Thus, blindsighting manifested by patient with damage to primary visual cortex (Pöppel, Held, & Frost, 1973, p. 295; Weiskrantz, 1986; 1990, pp. 247-278), covert recognition of faces demonstrated by patients with prosopagnosia, an impairment in face recognition (Bauer & Trobe, 1984, pp. 39-46; Damasio, 1990, pp. 287-296) or even unconscious perception occurring in patient with posterior parietal damage responsible for disorder of spatial attention, a pathology often called 'neglect' (Farah & Wallace, 1991, pp. 313-334), are largely reported and systematically observed symptoms. Each of these cases is a rebuttal of the widespread intuition stemming from folk psychology which suggests that one has not perceived anything unless one is consciously aware of what one perceives.[4] Undermining such commonsensical intuitions helps greatly to overcome part of the reluctance that some might express when faced with the project of treating intentional content independently from any considerations about conscious experience.

These remarks are not offered as a full justification for treating intentionality in isolation from the question of consciousness as it will be done in the course of this work. Providing such a complete justification falls outside the scope of this dissertation. I am just pointing to some reasonable ways to justify an approach that is shared by most naturalists. Some efforts needed to be made in that direction for this approach still

[4] One may entertain the thought that such experiments involve only mere sensory perceptions which do not count as cases of truly intentional contents. However, those unconscious perceptions do trigger overt intentional behaviors. For a technical yet accessible description and comments on those experimental protocols see *Visual Perception and Visual Awareness after Brain Damage: A Tutorial Overview* (Farah, 1999, pp. 203-236).

encounters limited but resolute opposition. Objections are not only coming from philosophers who tend to regard constraints put on their model by empirical sciences as dispensable if overridden by the necessary strength of some a priori metaphysical considerations. They are also voiced by some naturalist philosophers.

Jones Searle's position (1992; 1997) best exemplifies this radical opposition to the project of dissociating intentionality from consciousness. His model of mental representation depends entirely on the assumption that a physical state of the brain can be considered a mental state if and only if such a state is actually or potentially conscious. By establishing an intrinsic connection between the mental activity of the brain and the conscious life of the mind, Searle sets firmly on the side of common-sense intuition. Whether or not, by adding the proviso that being potentially and not necessarily actually conscious suffices for a state to be mental, Searle ultimately succeeds in offering a view consistent with the way contemporary science pictures our mental processes remains arguable.

Yet, Searle's position is motivated by a central philosophical argument which has proved compelling to many readers. Searle argues that by treating intentional representation in terms of mere computation, prevailing contemporary models (mostly functionalist ones) wrongly assume that semantics could be generated by algorithms processing information on the basis of purely syntactic rules. The failure to realize that semantics cannot be reduced to syntax is further explained by Searle as resulting from confusing true intentional states of the kind occurring in the conscious mind of living creatures with the kind he has dubbed "derived" or "as if" intentional states. These latter states are the ones conferred by human beings to artifacts that have been designed for simulating intentional behavior but do not have any mental activity on their own. Hence genuine intentional states when properly distinguished from derived ones do depend, according to Searle, on the existence of consciousness and cannot be properly understood without reference to it.

Searle's criticism are mainly directed to computational-functionalist models that is the kind of dominant models I will reject in this work in favor of a teleosemantic model based on Darwinian evolution. Searle has little to say for or against teleosemantics and the kind of criticisms he offers against standard functionalist models do not carry straightforwardly to teleosemantic approaches. Nonetheless the kind of models I am supporting shares with functionalist approaches the fundamental assumption that the question of intentional content can be and indeed must be treated independently of the one of consciousness. Let me just remark to conclude that one collateral benefit of the teleological approach I am advocating is to provide an analysis of the phenomenon improperly captured by the distinction between original and derived intentionality that, contrary to Searle's, is fully compatible with a separate treatment of intentionality from consciousness.

CHAPTER 1: THE DOMINANT MODELS AND THE CHALLENGE OF MISREPRESENTATION

The project of naturalizing mental content, that is of explaining how wholly physical organisms manage to represent the external world to themselves, has developed into two dominant strategies. Causal-informational theories such as the ones of Dretske (1981; 1988) or Fodor (1987; 1990) compete with computational-functionalist approaches like, for example, the Conceptual Role Semantics of Field (1977), Harman (1987) or Greenberg & Harman (2005). Causal-informational theories understand mental content in terms of covariation relations between mental states and the physical events in the external world responsible for causing such states. The basic idea is that the meaning of a given mental token must be traced back to the state of affairs that causes it to occur. On the other hand, functional-role approaches explain mental content in terms of the specific functional role that each mental state plays in the economy of the mental life of an organism in interacting with other mental states. Although based on distinct premises, these views can be, and indeed often are combined into elaborate mix models where they complement each other.

With regard to these dominant approaches, teleological theories such as the ones of Papineau (1984; 1987) or Millikan (1984; 1993; 2004), in which mental content is understood by reference to the etiological functions of representational mechanisms shaped by the selective pressure of Darwinian evolution have remained somewhat marginal and certainly less influential overall.

This chapter offers a criticism of the dominant models, providing some justification for reconsidering the actual value of teleological approaches to mental representation. In particular, it aims at demonstrating the inability of causal-informational models to overcome the problem of misrepresentation.

Section 1.1, *The Appeal of the Dominants Models*, accounts for the popularity of the non-teleogical perspectives in contemporary research. It presents a set of seemingly compelling reasons for thinking that the cluster of notions of 'nomologic relations', 'informational computation', and 'functional role' which are at the core of such perspectives could successfully break the 'intentional circle' and make possible a rigorous analysis of mental content.

Section 1.2, *Misrepresentation and Asymmetrical Dependency,* introduces the so-called 'problem of error' in relation to Dretske's indicator semantics, and shows how causal models fail in accounting for misrepresentations. Fodor's 'Asymmetric Dependence Theory', the most discussed and, apparently, the most compelling way to overcome the challenge of misrepresentation in the context of such theories is then introduced and analyzed.

Section 1.3, *Nomic Correlations versus Robustness*, explains why Fodor's strategy cannot succeed, at least not within the limits of the naturalist perspective. The failure of the 'Asymmetric Dependence Theory' helps to properly diagnose the problem of misrepresentation as a symptom of the fundamental inadequacy of the dominant models. Causal-informational models, I argue, are supported by two ideas. The first idea, explicitly endorsed by causal theorists, is the belief that the distinction between nomic correlations and contingent ones could account for the distinction between true representations and misrepresentations. The second idea remains implicit and is uncritically endorsed by causal theorists: the conviction that the phenomenon of misrepresentation can be reduced to a process of misperception broadly conceived. In this section, I introduce the notion of *robustness* of the mental content in order to challenge the soundness of these two tenets of causal-informational models.

In section 1.4, *Misperception versus Misidentification*, I argue that causal/informational models of mental content overlook the normative element attached to intentional representations and wrongly focus on misperception rather than misidentification as the explanation for misrepresentation.

1.1 The Appeal of the Dominant Models

Several reasons converge to explain why most naturalists favor causal-informational models of intentionality. It is common practice in folk psychology to explain the intentional content of a thought by making reference to the background beliefs of the person having such a thought. Those beliefs are in turn interpreted as responsible for the propositional attitude of which the content of this thought is a constitutive element. Hence, Rodrigue's decision to challenge the Comte de Gormas in a duel is explained by Rodrigue's desire to restore the honor of his family, a desire that is explained itself by the belief that his father has been humiliated by the Comte. The explanatory and predictive powers of such an interpretative strategy in regard to human behavior are not to be underestimated, as they are continuously validated by the successful use of folk psychology in daily life.

Yet, this kind of explanation is problematic on many accounts. To start with, such an elucidation of the mental content of a given thought, by reference to its specific location within the network of beliefs and desires held by its owner, depends itself on an intentional reading of this background network of mental states. Naturalism needs to break this ‘intentional circle’. Mental processes are part of the natural world and the explanation for the production of thoughts must be continuous with the type of explanation that science is offering for other kinds of natural phenomena. Ultimately, one needs to provide an explanation for mental content and intentional behavior on the basis of entirely non-intentional processes and entities. Allowing for a free-floating realm of mental activities entirely autonomous from any physical-biological mechanisms would amount to conceding dualism.

Such considerations have led some philosophers, like Paul Churchland (1984) and Patricia Smith Churchland (1986), to reject folk psychology altogether. They regard it as a misleading theory of the mind that needs to be radically revised, if not entirely discarded. According to them, it is the neuro-biological approach to cognitive processes that offers the only legitimate ground for philosophy of mind. Everyday notions of ‘beliefs’, ‘desire’, ‘will’, ‘attention’, or ‘memory’ are confused and ill-conceived. At best, these notions offer a crude and naïve picture of clusters of highly complex neurophysical mechanisms which should be distinguished and investigated. In the worst case, they may simply not refer at all. Hence, in the near future of neuro-philosophy that the Churchlands are advocating, some common notions in folk psychology might have no more epistemological value or ontological import than phlogiston has in contemporary physics.

However, naturalist philosophers in general have resisted the call for eliminativism. Most of them see themselves as intentional realists. They defend the view that, as real phenomena, beliefs and desires truly have causal powers; furthermore, they have such powers to the extent that they are beliefs and desires. It is unclear how eliminativism could be compatible with the Davidsonian claim that reasons are causes. Besides, as Jerry Fodor (2008) is fond of pointing out, neuropsychology textbooks are replete with explanations referring to memory, visual perception, and other mental processes described in folk psychological terms with the attempt to identify the brain mechanisms responsible for each of these mental faculties. Thus, there seems to be a self-defeating element in the project of explaining intentional psychology in terms of neuropsychology while simultaneously claiming that neuropsychology itself depends on the dispersion of folk psychological notions.

That being said, the naturalist inclined to reject eliminativism in favor of a milder kind of reductionism still needs to break the intentional circle by tracing the emergence of mental processes back to their underlying physical properties. This is what the causal element in causal-informational models is supposed to guarantee.

Physicalism and behaviorism have competed in offering two opposed ways to elucidate such a causal link. While physicalism was aiming at identifying mental tokens of a given type with their physical counterparts in the brain, behaviorism was engaged in reinterpreting the mental activity of an organism in terms of its actual or potential dispositions as manifested in overt reactions to external stimuli. The respective limitations of these divergent approaches quickly became apparent. Ned Block (2008) has coined a suggestive terminology that helps to sum up the issue. Hence, following

Block, one might call type-physicalism too ‘chauvinistic’. By equating any given type of mental states with its actual type of physical realization, physicalism prevents two organisms with different physical make-ups (for example, members of two different species) to share any common thoughts. Behaviorism, on the other hand appears to be too ‘liberal’ by endowing any automaton capable of duplicating some well-specified behavior with a mental life of its own.

By identifying thoughts with abstract functions of the system, functionalism offers the possibility for multiple realizations of mental states. These states can be actualized by different mechanisms which can themselves be implemented on the basis of different physical supports. In addition, and contrary to behaviorism, functionalism allows internal states, notably other mental states, to play a decisive part in explaining the functional role and thus the causal power of any given thought. As a result, two physically different organisms using distinct mechanisms for the production of intentional content, yet implementing a common abstract function, can be said to entertain the same thought. These reasons explain why most of contemporary causal-informational models are included into broader functionalist theories of the mind.

The informational component in these models comes in response to the second issue raised by the everyday practice of describing mental states. Such states are often presented by folk psychology as incorrigible to the extent that they are subjectively experienced states, directly accessible only from a first-person point of view. These are qualitative states, some so rich, complex, or idiosyncratic that they are said to be ineffable. Science, on the other hand, requires theoretical claims to be public, unambiguous and transmissible. Scientific claims are supposed to quantify over well-defined and measurable objective phenomena, accessible from a third-person point of view.

Such considerations certainly account for the tendency of many naturalists to treat intentional content as information and the production of representations as information processing. Implicit analogies with successful reduction that have occurred in the past appear to be important influences here. There is little doubt that paradigmatic cases like the one of thermodynamics are in the mind of many naturalists. They seem to hope that representation could eventually be identified with the activation of causal-informational paths the way heat has been identified with molecular motion. In addition, a well-defined notion of content in terms of information could be substituted for the subjective notion in folk psychology, the way the precise gradients of temperature on a thermometer are to be preferred to the subjective and relatively vague reported sensations of warmth and coldness.[5]

Finally, the emergence and rapid development of artificial intelligence has contributed to the reinforcement of this general move toward functionalist causal-informational models by giving weight to the idea that the mind could be to the brain what the software is to the hardware that runs it. This idea captures in one unified picture all the key elements of the dominant perspective: the establishment of nomic connections, the reduction of information to discrete quantified units, the multi-realizations and the various implementations of the function responsible for producing a common thought in

[5] The mathematical theory of information inaugurated by Shannon's work (1948) has provided the conceptual basis for such a rigorous treatment of the notion of information, while the widespread practice of downloading, compressing or transferring data on computers has eventually rendered this formalized approach familiar and almost intuitive to an ever-growing number of non-experts.

different systems. The convergence of these different elements helps in understanding the predominance in contemporary naturalist philosophy of mind of theoretical models best described as causal-informational/computational models of mental representation.

1.2 Misrepresentation and Asymmetrical Dependency

Hence, non-teleological attempts to naturalize mental content are generally conducted in terms of causal-informational/computational approaches. The way the challenge of misrepresentation threatens such models may best be seen in the case of Dretske's indicator semantics. The core idea of indicator semantics is to explain the representational properties of a particular token by the fact that such a token belongs to a type nomically connected to a given type of objects or state of affairs. Representation occurs thanks to the information carried by this token, which serves as a reliable indicator that a certain situation holds in the external world. Hence, the token 'dog' means DOG and refers to dogs by being usually present to the mind if and when—although not only if or necessarily when—the actual presence of dogs is experienced. In this sense, the token 'dog' means DOG as a result of being an indicator of dogs just as the speedometer's hand pointing at '35' indicates that the car is moving at 35 mph.

However, indicator semantics seems unable to make room for the occurrence of errors. Thus, if we suppose that 'dog' means DOG (i.e. that dogs and only dogs in its extension) because it's a law that dogs cause 'dogs' then we face an alternative: either only dogs in fact really cause 'dogs' and errors should never occur, or some non-dogs, such as wolves in the dark, can cause 'dogs' as well. But then, if symbols express the properties whose instantiations are nomically sufficient for their tokening, one must conclude that ‘dog’ expresses the property of being either a dog or a wolf in the dark, and therefore means DOG and/or WOLF IN THE DARK. The conclusion is that error is impossible.

Among the different proposed ways out since the problem has been acknowledged, Jerry Fodor’s deserves particular attention. The reason why his solution remains unsatisfactory helps to better appreciate how the challenge of misrepresentation is just a symptom of the inadequacy of the philosophical understanding of the nature of intentional content in which causal/functionalist models for mental representations are rooted.

Fodor’s idea (1987; 1990) is that an ‘asymmetrical dependence’ exists that permits discrimination of accurate representations from misleading ones. An actual law is in place that connects the adequate type of mental or linguistic tokens to their references, in our example a law connecting ‘dog’ to dogs, on which the accidental and derivative connection between ‘dog’ and wolves depends, while it is not the case the other way around. Intuitively the idea is that false representations depend on true ones. It is only because dogs are indeed nomically responsible for generating the mental token ‘dog’ that it is possible to mistakenly generate such a token in the presence of a wolf that looks like a dog. By contrast, no reference to any causal connection to wolves is required to explain the production of the mental or verbal token ‘dog’ in the actual presence of a dog and therefore no reference to any wolf is required to explain why ‘dog’ means DOG.

There is an air of straightforwardness and elegance in such a solution. However complications of all sorts arise as soon as one tries to explain in details how exactly this solution is supposed to work. Over the years Fodor has been induced to perpetually return to the Asymmetric Dependence Theory (ADT hereafter) working hard to clarify his position, sometimes revising and complicating it, sometimes amending it while considering no less than eleven different lines of possible criticism (Fodor, 1994, pp. 88-136). Giving an accurate rendering of Fodor's theory is made difficult by these constant modifications and adjustments.[6] The real issue with regard to the present work is not so much to decide whether Fodor's effort could be successful or not, but rather to wonder, were it be successful, whether or not ADT explanations will remain naturalist explanations.

Fodor's treatment of the challenge of misrepresentation as faced by causal-informational theories of content is, I believe, threefold. First is introduced what Fodor dubs the Crude Causal Theory of content (CCT here after), i.e. a basic account for mental representation that is supposed to capture the common core of causal-informational theories. Secondly, CCT is shown to be unable to make room for misrepresentation; that is the so-called problem of error already analyzed before in relation to Dretske's model. CCT helps to better understand the actual nature of this problem which turns out to be only the most striking aspect of a larger issue, namely the "disjunctive problem". According to Fodor the disjunctive problem is germane to any theory of content that makes use of CCT as its basic structure and that is any causal theory of content. Third,

[6] Millikan (1992), for example, seems to think that providing such an account is nearly impossible because of the obscurity of a theory that keeps changing all the time. In her review of Fodor's Theory of Content 2 (TOC2 in the following quote) she writes: "I must confess no luck from the start (Psychosemantics) in determining exactly what the AD thesis is. I cannot wade far, I can just flash the light into the dark waters of TOC2 to glimpse how the target moves" (p. 900) and "the rules of the game change quite systematically every time someone passes GO" (p. 901).

Fodor offers ADT as a solution to the disjunctive problem. Let us now follow these three steps. First CCT is presented as the view according to which:

> The symbol tokenings denote their causes, and the symbol types express the property whose instantiations reliably cause their tokenings. So, in the paradigm case, my utterance of 'horse' says *of* a horse that it *is* one. 'Reliable causation' requires that causal dependence of the tokening of the symbol upon the instancing of the corresponding property be counterfactual supporting: either instances of the property actually do cause tokenings of the symbol, or instances of the property would cause tokenings of the symbol *were they occur*, or both. I suppose that it is necessary and sufficient for such reliable causation that there be a nomological—lawful—relation between certain (higher-order) properties of events; in the present case, between the property of being an instance of the property *horse* and the property of being a tokening of the symbol 'horse.' The intuition that underlies the Crude Casual Theory is that the semantic interpretations of mental symbols are determined by, and only by, such nomological relations. (Fodor, 1998, p. 99)

On closer inspection, the rendering of CCT offered here is somewhat misleading. While providing a suggestive illustration, the utterance of the word 'horse' in the presence of the actual animal is hardly a paradigmatic case for which the theory is the more likely to offer a successful explanation. First and foremost CCT is to apply to the tokening of mental symbols and not so readily to the utterance of linguistic expressions. This is so mainly because, as Fodor acknowledges himself, Gricean assumptions and other pragmatic considerations about context relevance make the causal dependence of tokenings of English expressions upon semantically relevant wordy situations typically very complex and indirect. By comparison, causal links connecting mental representations to their own relevant wordy set ups are generally more straightforward. One should therefore expect the latter connections to be more reliable than the former.

The horse example turns out to be misleading in another way as well. For in fact, CCT demands that the tokening of a mental symbol 'horse' be not referring to the presence of a horse, but rather be denoting the existence of the instantiation of the property of being a horse by a given object present in surroundings. Hence Fodor's rather contrived statement: "my utterance of 'horse' says *of* a horse that it *is* one". We will return to the reason why exactly according to Fodor, it has to be so, later on in this chapter. The point to be stressed here is that in the quote above, Fodor presents an account of CCT in terms of linguistic expressions denoting the objects responsible for their utterance, a view both intuitive and appealing, but ultimately inaccurate.

A more adequate account would picture CCT as a theory capable, in the best scenario, of explaining how the occurrence of mental tokens of a given type can be triggered by the instantiation of the property of the kind responsible for making its bearer the particular thing, or rather, the particular kind of thing that it is. When stated accurately, that is in strict compliance with its complete list of specifications, CCT looses most of its intuitive appeal. One may wonder what is this unique property that makes a horse a horse or at least that commands an observer to think that he is facing one. Some have suggested ‘looking like a horse’ or ‘looking horsey’ as the best candidate for being such property. This suggestion brings with itself a host of issues requiring clarification, to be discussed later on in this section.

At this early stage, I should content myself with making the two following remarks. First, horses look like horses because they are horses and not the other way around. Second, there is a sense in which any given thing always looks the way it does because of what it is, however unusual the conditions of its manifestation turn out to be and however different its appearance may be when compared with the way it looks under standard circumstances. Should the ADT theorist be willing then to continue his pursuit by engaging in the cumbersome task of stating the conditions under which things could be safely assumed to be looking as having the property of being what they are?

That being as it may, the reason why CCT is viewed by Fodor as a crude theory stems from different considerations altogether. The basic idea of causal theories being that the meaning of a mental token is whatever causes this mental token to occur, CCT must have it that ‘horse’ means ‘horse’ because it is *only* for horses and for *all* horses to produce the occurrence of the token ‘horse’.

Both the ‘all’ clause and the ‘only’ clause are clearly unsupported by empirical observations. It is not the case that every horse will cause the tokenings of ‘horse’ for some horses may be out of reach of any observer; others may remain undetectable while being in the surroundings because of darkness or visual obstacles. It is only when the conditions are satisfied for a proper observation from the observer point of view that horses may be expected to generate ‘horse’ tokenings. That is why the ‘all’ clause should be restated in terms of counterfactual statements in which it is assumed that the horses failing to trigger ‘horse’ tokenings would have indeed triggered such tokenings, had the conditions been proper. To state such conditionals in naturalistic terms, without begging the question, could be expected to be a challenging task. We will see later on how and to what extent, Fodor manages to provide such specifications.

Meanwhile, the ‘only’ clause presents yet another challenge, for it is easy to imagine situations in which a particular object that is not a horse would nonetheless trigger the occurrence of the token ‘horse’. This leads us back to the problem of error. We have seen before how Dretske’s indicator theory was making misrepresentation impossible. Fodor’s CCT helps to see how the challenge of misrepresentation is a general challenge faced by any causal theory and how such a challenge can be restated in terms of the ‘disjunctive problem’. Thus ‘horse’ tokenings are generally horse-caused but under untypical conditions ‘horse’ tokenings can be also cow-caused. It is tempting to dismiss cow-caused tokenings of ‘horse’ as misrepresentation. However, if the meaning of a token is what causes its occurrence, then it seems that, being sometimes caused by a cow, ‘horse’ has to mean COW at least in some circumstances.

Conversely, we may find cases where different causal paths could be competing in triggering the occurrence of symbol tokenings of a given type S so that each token ‘S’ could be indifferently nomologically A-caused or B-caused. In this case, the content of ‘S’ will be expressing the disjunctive property of being (A or B). CCT therefore does not have the resources for distinguishing a case of misrepresentation from a case of disjunctive representation.

Because it affects CCT, and CCT is at the core of any causal theory of mental representation, the disjunctive problem permeates all different versions of a causal account for mental representation. On the other hand, if a solution can be offered to correct this problem for CCT, one can hope that it will carry on to any version of causal representation. Fodor's ADT is supposed to offer such a solution.

It follows from Fodor's analysis that the identification of misrepresentation is made impossible because of a symmetry between A-caused and B-caused tokenings of a particular symbol 'S', making nomological connections difficult to tell apart from contingent ones. Fodor's ADT purposes to break such symmetry by appealing to the difference between the counterfactual properties of these causal relations. To repeat, it is only because, according to Fodor, there is a nomic relation between 'horse' tokenings and instantiation of actual horses that the misidentification of a cow for a horse can sometimes lead to the tokening of 'horse'. By contrast, no connection between cows and 'horse' is needed to explain why horses cause 'horse' tokenings. Precisely stated, Fodor's ADT reads as:

> B-caused 'A' tokens are wild only if the nomic dependence of instantiations of the property of being an 'A' tokening upon instantiations of the property of being a B tokening is itself dependent upon the nomic dependence of the property of being an 'A' tokening upon instantiations of some property other than B. (Fodor, 1998, p. 164)

We can now understand why Fodor wants his view to be stated in terms of nomic dependences between higher-order properties, namely the property of being a particular kind of object and the property of being a token of a particular kind of symbol, rather than simply between objects and symbols. The reason can be seen by reflecting on the following challenge offered to Fodor by Scott Weinstein. Hence, consider the situation bellow:

> (i) Small horses cause 'horses.'
>
> (ii) Horses cause 'horses.'
>
> (iii) (i) depends on (ii) (small horses wouldn't cause 'horses' unless horses did).
>
> (iv) (ii) is not dependent on (i) (horses would cause 'horses' even if small horse didn't; even if, for example, there were only large horses).
>
> (v) So small-horse-caused 'horse' tokenings are asymmetrically dependent on horse-caused 'horse' tokenings; so small-horse-caused horse tokenings are wild. (Fodor, 1998, p. 164)

Hence, as stated, CCT leads to the unwelcome conclusion that small-horse-caused tokenings of 'horse' amount to misrepresentations. Fodor offers to avoid this unfortunate conclusion with the help of a 'parade version' of the argument in which premise (i) is replaced by (Pi):

> (Pi) 'Horse' tokenings are nomically dependent on the instantiation of small horse.

The conclusion that asymmetric dependence is not sufficient for wildness is then avoided by acknowledging (Pi) to be false on the account that the counterfactual supporting connection is between ‘horse’ tokenings and horse instantiations not between ‘horse’ tokenings and small horse instantiations.

The justification offered here is that the small horse responsible for the occurrence of the tokening ‘horse’ would have caused this very same tokening to occur even if it had been larger. Fodor’s parade version is an attempt to recourse to the higher-order property of being a horse, a property that manifests itself by a horsey appearance, to rescue the explanatory power of ADT.

As such though, this attempt is clearly unsuccessful. Fodor wants to argue that since a small horse would have caused a ‘horse’ tokening even if it had been larger, the nomic connection has to be between horses and ‘horse’ tokenings rather than between small horses and ‘horse’ tokenings. Hence, Fodor appeals to carefully chosen counterfactuals to reach the necessary conclusion that the connection between small horses and ‘horse’ tokenings is contingent and asymmetrically dependent on the nomic connection between horses and ‘horse’ tokenings. Fodor seems to fail to realize that a similar strategy can be used to challenge the claim that the relevant nomic connection is between ‘horse’ tokenings and horse instantiations (that is the instantiation of the property of being a horse). This is so, because one could point out that each of the horses responsible for those ‘horse’ tokenings would have caused a ‘horse’ tokening even if it had been a cow, providing that this cow looked horsey enough.

1.3 Nomic Correlations versus Robustness

It is a telling fact that a commentator like Hans-Robert Cram (1992), who has come to realize that the parade version of ADT does not successfully solve the disjunctive problem, nonetheless praises Fodor's for giving a central role to higher-order properties in accounting for the nomological connections underpinning the production of intentional content. Cram's diagnosis of the failure of the parade version is that Fodor's specification of ADT remains incomplete. Hence, with the help of the parade version, ADT explains the wildness of B-caused 'A' tokenings by the fact that there would be a causal route from A's to 'A' even if there were no such route from B's to 'A' while the reverse does not hold. Cram subscribes to such an explanation but complains that no reason is presented by Fodor for why such a counterfactual relation holds in the first place. In short, Fodor does not explain why it is that counterfactual-situations have the precise features needed to guarantee the right semantic discrimination resulting from causal nomic connections between kinds of objects and kinds of tokenings.

Cram's complaint is legitimate here. Yet in supporting Fodor's parade version, presenting ADT as a partial theory waiting to be properly completed, Cram sides with the long list of commentators, proponents and critics of Fodor's model alike, who fail to properly identify the real nature of the challenge of misrepresentation. Cram's following remarks are in that sense typical:

> The representation relation is not a relation between symbol types and objects, but rather, as stated in the 'parade version', a relation between symbol types and properties; in Fodor's example, the symbol 'S' is not

> supposed to represent horses, but rather the property of being a horse. And I want to add that the symbol represents this property only in virtue of another property horses usually have, namely the property to look like a horse. Indeed, the problem of misrepresentation consists precisely in the fact that not only horses, but sometimes also cows have the property of looking like a horse and can therefore cause the symbol 'S' via this property (Cram, 1992, pp. 56-72).

One can imagine why the appeal to the higher-order property of being a specific kind of thing is regarded by Cram as an improvement upon the basic structure of CCT. Thus, one may entertain the intuitive idea that misrepresentations are generally sporadic events resulting from accidental misperceptions occurring under odd circumstances. Misrepresentations so understood are unlikely to resist a closer examination of the actual states of affairs causally responsible for their production. Under uncooperative circumstances, an object will feel, smell or look as being something else, but it is very unlikely for the circumstances to be such that it will deceive all our senses at once in a coherent and sustainable manner as we keep triangulating the information coming from independent sensory paths.

However, misrepresentation does not reduce to misperception even in the basic case of the identification of a physical object through direct sensory perceptions. Contrary to what Cram claims, it is simply not true that "the problem of misrepresentation consists precisely in the fact that not only horses, but sometimes also cows have the property of looking like a horse and can therefore cause the symbol ‘S’ via this property”.

Notice first the ambiguity of the expression ‘looking like a horse’ in such a context. If the assumption is that misrepresentations are indeed misperceptions to be dispelled by a process of triangulation of our sensory perceptions, than ‘looking like a horse’ should read as ‘presenting the external aspect’ of a horse. There can be no legitimate reasons for privileging seeing over touching, smelling, and so on. In fact it is precisely under the hypothesis that the higher-order property of being a horse generally manifests itself through different modes of appearance, each soliciting different senses, that one can be tempted to understand the process or mental representation as the one of properly identifying this higher-order property. Misrepresentation, following this misleading line of reasoning, reduces to misperception, and misperception is in turn explained by the fact that the observer has only a limited access to the full array of modes of manifestation by which this higher-property signals its presence under normal circumstances.

While it may be true that misrepresentation is the exception because wild tokenings are typically rare, it must be observed that, conversely, accurate representation is the common law only because normal tokenings are typically robust. Robustness is a technical notion in use in different scientific fields. In computer sciences, for example, an algorithm is declared robust if capable of keeping processing information properly when the computer is fed with abnormal inputs. In biology an organism phenotype is robust when it remains constant in spite of genetic mutation.

In the context of our discussion, the notion of robustness refers to the fact that a given item can maintain its particular meaning while occurring in a great number of very different circumstances and through very different modes of manifestation. Such a property manifests itself continuously in everyday life. Some people will want to reassure themselves of the exceptional quality of the wine they plan to serve to their guests by tasting it, others with a less educated palate will want to check the price on the bottle. On the other hand, think about what happens with two-dimensional animated cartoon characters. Tex Avery's characters, Droopy and The Wolf, are spontaneously recognized for what they are, namely a dog and a wolf. They seldom generate misrepresentations even among the young kids who constitute their main audience despite the fact that they possess many attributes in common with humans and share only very few with their natural models. The property of robustness is also supporting the use of linguistic items which manage to convey a stable meaning whether they are voiced, hand-written, or printed. Hence, the accurate tokening of a particular type of token is typically caused by very different wordy setups, some not even always figuring the actual object that the token is properly representing for the occasion.

It is not only that the phenomenon of robustness sheds some suspicions on the idea that misrepresentation could be reduced to misperception; robustness also questions the assumption, present in causal/indicator theories, according to which the distinction between nomic causal relations and contingent ones plays a central role in accounting for the production of erroneous mental content. This last point requires careful examination.

The causal theorist has it that the mental token 'horse' gets its content by being causally triggered in the presence of horsey looking items. With the help of the parade version, ADT states that in the standard case, the higher-order property of belonging to the horse kind explains the animal's looks and this look is causally responsible for triggering mental tokens of the 'horse' type in the observer's mind. Causal theorists, then, explain misrepresentation by the fact that the causal link between looking horsey and being a horse is not always reliable. Horses do not always look horsey and cows sometimes do.

That being said, if for a mental token, meaning HORSE is having the higher-property of being a token belonging to the 'horse' type, the expression 'horse type' cannot merely be naming the set of all the mental tokens generated by horsey looking things, because then, the extension of the mental type HORSE, and therefore the meaning of any 'horse' tokening, will be disjunctive. This is supposed to be corrected by the fact that only horsey features *properly generated* count, the other ones resulting in misrepresentations.

To remain fully naturalistic, causal/indicator theorists needed to unpack the notion of being 'properly generated' in strictly non-intentionalist terms. Their strategy has been to heavily rely on the distinction between nomic/necessary versus accidental/contingent causal connections. Hence, Dretske's original account of informational semantics describes the kind of natural information responsible for generating accurate mental representations as one carried by its informational channel from its source to its indicative signal in accordance to some natural law. Misrepresentation is explained by the unreliability of signals resulting from the accidental occurrence of some contingent connections between an indicator and its source. Thus the basic idea is that the distinction between nomic connections and accidental ones, or mere frequencies, accounts for the difference between true representations and errors.

On the one hand is the necessary nature of the causal connection between the source and the signal when information is carried by a natural-law-governed channel. On the other hand is the univocal nature of the semantic relation between the intentional content of a given type of mental tokens and the type of things it is truly representing when the act of representation is successful. The indicator semanticist would like the latter to be a straightforward reflection of the former. The same can be said in the case of standard causal theories. Helped with the parade version of ADT, the causal theorist understands misrepresentation as misperception of types, and pictures nomic connections as holding between kinds of items and types of mental tokens. The precise content of the semantic relation between a mental token and its reference is then claimed to be a direct reflection of the higher-property of belonging to a given type.

The causal theorist's idea seems to be that some necessary (most likely intrinsic) properties identify any particular object as a member of its kind. Such properties manifest themselves by being causally responsible for the overt features which the object presents to the observer. Those features are in turn responsible for triggering the occurrence of mental tokens of the right kind in ordinary setups.

However, robustness shows the list of the potential features susceptible of being involved in the production of the same accurate representation to be impressively long and diverse. There is no reason to imagine that the connection responsible for the identification of a referent as being an instantiation of a particular type has to be the same for an observer from one occasion to another. And certainly there is no reason to think that it is the same from one species to another. Thus, there is no reason to assume that a single nomic relation holds between the higher-property of being of a given kind, assuming such a property exists in the first place, and the proper type of mental tokenings representing that property.

Furthermore, the causal/informational strategy seems to rely on the ontological assumption that states of affairs in the external world are made up of objects naturally ordered within distinctive types, according to some necessary features that they possess as a result of binding natural laws. In short, such accounts of mental representation, based on nomic correlation between kinds of items in the world and types of mental tokens in the head, seem to postulate the existence of entities with a common higher-property that make them all members of a given set or class. Remember Fodor's claim that "in the paradigm case, my utterance of horse' says *of* a horse that it *is* one." Hence, properly identifying a horse would amount to recognizing that a particular object has the property of belonging to the horse kind. However, this property itself cannot possibly be identical with any one of the overt physical features of the animal to be identified, such as its size, the color of its coat, or the number of its legs. None of these features are exclusive to horses nor are strictly required for an animal to be a horse. The property of belonging to the horse kind must therefore be understood, in this context, as the higher-property of possessing the set of necessary properties shared by all the members of the kind horse, understood as a given class of entities.

The appeal to such a higher-property is problematic in two ways. The 'looking as a horse' property, now understood as the higher-order property of belonging to the horse kind, could be said intrinsic to horses, and wrongly ascribed to cows looking like horses, only if one has already secured an independent path for identifying the higher-property that signals horse-kind quite aside from the ordinary way it reveals itself through the observable features of horses and horsey-looking cows alike. As noticed earlier, this in turn, would require that the counterfactual conditions under which a thing is guaranteed to reliably present the appearance of having the property of being what it is be already established; a cumbersome task at the end of which, if it is ever to be completed, lurks the risk of an infinite regress.

Problematic also, is the type to type-nomological relation that both supports and mediates the representational relation between items and their mental tokenings as presented by causal/informational theorists. It would follow from their models that the standard way of identifying an item would first require subsuming it under its particular natural kind of entities. Here again, the unjustified tendency to reduce misrepresentation to misperception is implicitly at work. Hence we can imagine that looking at a person coming from the other side of the street, I will easily identify her as a human being before she is close enough for me to identify her as a woman. I will then need for her to come even closer before I am able to put a name on her face and identify her with some confidence as a particular individual.

However one should notice that the identification of an individual is not generally mediated by the pre-identification of its kind nor it is the case that the identification of an individual is in general more difficult to achieve than the identification of the kind to which this individual belongs.

A recent trend in the US has been for women not disposed to conceive but willing to experience a mother-child-like bound to adopt young monkeys and treat them like human babies. The behavior of these young animals gives little room for doubting that each of them identifies its care taker as a specific individual. Yet there is no reason to think that they suspect this individual of being a human being and not a monkey. The process of identification as practiced on a daily basis by living organisms, humans included, appears to be largely insensitive to the ontological hierarchy that causal theorists present as playing a key role in fixing the intentional content of mental representation.

The ease with which individuals are commonly identified as such, challenges also the idea that the distinction between nomic connections and accidental causation is responsible for the distinction between proper representation and misrepresentation. For it is for properties and not for individuals to be subjected to nomic generalization. This is reflected in the well-known fact that there is no such thing as the science of a singular object.

Hence, even if Dretske's indicator semantics were to be correct, its scope would remain restricted to the explanation of how and under which conditions an organism could properly represent to itself that x is F and wrongly represent that x is G when (Fx) is in fact the case. But neither Dretske's indicator semantics nor any other causal model based on nomic correlations seem, as such, to have the resources for explaining the fact that we are able to identify the particular object which in each occasion satisfies (Fx). Our ability to commonly individuate individuals remains unexplained.

I would like to suggest a possible way by which the causal/indicator theorist could attempt to respond to the challenge of the identification of individuals. The solution I will present now is, as far as I can tell, the best possible move in the context of such theories, and yet it remains deficient. Reflecting on the reasons why it is deficient will help to improve our diagnosis of the general inadequacy of causal/informational models.

Hence, here is a (ultimately unsatisfying) proposal about how individuals could be identified in the context of theories of mental representations based on nomic correlations. First, x is properly recognized as having the higher-property of belonging to a given kind of entities. This is done, one supposes, by rightly attributing to x the property (H), a property that x actually possesses as a member of its kind in virtue of a law governing all the members of such a kind. Secondly, another property, the property (I), is rightly attributed to x. Here, x truly possesses the property (I), but not as the result of being a member of its kind. In itself, the identification of this second property (I) of x will not be enough to identify the particular individual standing for x, for contrary to (H), (I) is a contingent property that x is likely to share with many individuals belonging to other kinds as well. However the pre-established knowledge that (Hx) is the case restricts the extension of the set of variables that could possibly stand for what is represented. Within this restricted set, the contingent property (I) attached to x, precisely because it is not a shared property characteristic of the members of this restricted set, helps now to discriminate among them the individual that does possess such a property, leading to the identification of the individual that, in this particular occasion, satisfied (Hx).

We could illustrate this theory with our horse example. Hence, imagine that one morning you are visiting Bob's farm which possesses several pigs, cows, geese and yet other animals, among which are few horses. You ask Bob the name of one of the horses which captured your attention because of the peculiar black and white pattern of its coat that contrasts with the uniform color of the coat of the other horses. You learn from Bob that the horse's name is Domino. Later in the evening you return to the farm where you recognize Domino and you call it by its name. According to the model just presented your ability to identify x as being Domino, results from your ability to first identify x as being a horse, then to your ability to notice the black and white pattern of the horse's coat and from the conjunction of these independent premises to establish that x is Domino.

The natural law supporting the correlation between the horsey appearance of Domino and its belonging to the horse kind will not have been enough to identify Domino as Domino. On the other hand Domino is likely to share the black and white pattern of its coat with many of the cows on the farm. You may not remember the precise configuration of the spots on Domino's coat so as to distinguish it from alternative similar black and white patterns. In addition, displaying such a black and white pattern is not a property that can be subsumed under any natural law of the sort that supports a nomic relation between types of mental tokens and natural kinds of items. However, thanks to the natural law governing the horsey appearance of the members of the horse kind, the domain of the possible representations has been now safely restricted to horses. Within this restricted domain, the black and white pattern of the horse's coat because it is a contingent feature attached to a particular individual, rather than an essential property shared by all the members of the same natural kind, helps to properly identify Domino.

What then is wrong with the idea that the identification of Domino could be achieved thanks to the following deduction?

1) x has the property (H) of being a member of the horse kind

2) x has the property (I) of having a black and white coat

3) Domino is the black and white horse

C) Therefore x is Domino

An additional premise is clearly missing for the conclusion to be properly derived, namely the fact that the scene takes place on Bob's farms. The black and white pattern of the horse's coat is a reliable indicator of Domino's identity only because the set of potential candidates for being identified as possessing such a black and white coat is severely limited (in this case to only one horse) by the particular time and location in which your observation occurs.

More generally, the co-occurrence of a particular sign with the thing it indicates or signifies typically persists only within a well-delimited—or at least stable enough—domain. Your ability to infer the identity of the horse depends on your ability to properly track the local domain of occurrence of the sign supporting this identification. In the present case you are aware that you are back at the very same farm you visited this morning and that the short amount of time between your two visits makes it unlikely for Domino to have been sold and/or for Bob to have acquired another black and white horse.

Of course, none of these considerations need to be actually explicitly formulated nor consciously pondered upon when recognizing Domino. In fact, it is a mistake to imagine that successfully representing something generally pre-requires the ability to properly identify the conditions that need to be satisfied for the representation to be successful. Animals commonly identify natural signs of the presence of food, mate or predator in the surroundings and act accordingly. While the co-occurrence of such natural signs and the presence of the items that such signs indicate often depend on the very specific set up of the animal's territory, there is no need for the animal to have a mental representation of this fact. As long as sedentary habits, geographic barriers or limited mobility restrain the scope of the animal migrations, the territory of its living will persistently overlaps the proper domain of occurrence of the natural signs the animal relies on for representing its surroundings to itself, in a way that guaranties the satisfaction of the animal's needs. The animal's ability to have true representations of its surroundings does not depend on the animal's ability to have true representations of the set of necessary conditions needed for that to be the case.

Millikan (2004) who was one of the first to point out the inability of indicator semantics to account for the identification of individuals, has also criticized Dretske's notion of natural information precisely for being a context-free notion. She rightly insists that correlations between what, in her terminology, are called "local recurrent signs" and the things they represent need to be defined relatively to relevant reference classes of items, within the boundaries of some specified domain. Thus, the occurrence of tracks made by pheasants on the ground of a given area is a recurrent natural sign of pheasants in that area. The same tracks found in an area depleted of pheasants but inhabited by quails is a recurrent natural sign, hence, a reliable indicator, of the presence of quails, not pheasants. Also, the pheasants' tracks illustration should not lead one to conclude, nor, I take it, is Millikan's intend to imply, that the relevant context is always best specified in terms of geographic boundaries. In fact, even in the pheasants' case, the relevant context seems to be better described in spatiotemporal terms. More generally, I see no good justification for trying to decide a priori in favor of a limited list of pre-established ways to circumscribe, in each case, what constitutes the relevant context.

Fundamentally, Millikan's pheasant's tracks story is a suggestive illustration of a deeper and more abstract principle: no additional constraints put on the connection between signs and the things they indicate—whether they would touch on the necessary nature of the connection, the high frequency of the occurrences, or the proper conditions under which it must be observed—can ever be combined so as to provide a context-free account of the information that local recurrent signs are said to carry.

Furthermore, you took the black and white pattern of the horse's coat as a reliable indicator of the horse identity because you assumed that this pattern was likely to remain the same from one encounter to another with Domino. This morning Domino was the only black and white horse, but it was also the only horse in the field standing a few feet away from the big oak tree. Although you may have noticed Domino's position as well, you didn't take it to be a reliable sign for future identification. What seems to be needed for a sign to be a reliable indicator of the presence of an item or of a particular state of affairs is that the frequency of the co-occurrence of the sign with what it indicates reflects some persistent connection between the two. Tracking down such persistent connection and being able to guess which sign or type of signs is susceptible to serve as a reliable indicator in the future, seems to play a central role in our ability to form accurate mental representations.

Yet, notice that the color of Domino's coat proves to be a reliable indicator of Domino's identity despite the fact that the correlation between the presence of the black and white pattern and the presence of Domino is not subsumed under any natural law.

That is not to say, of course, that there is no causal explanation for why Domino's coat displays this particular pattern nor why this pattern remains attached to Domino's coat, hence to Domino itself, over time. In other words, the success of mental representation depends on the ability to detect projectable properties of the thing being represented rather than on subsuming the relation between this thing and the sign of its presence under a type to type correlation governed by some natural law.

I have called attention to the phenomenon of robustness in order to challenge what I take to be two of the basic tenets of causal-informational theories. The first one is explicitly endorsed by causal theorists and states that the distinction between nomic correlation and accidental causation accounts for the difference between true representation and misrepresentation. The second one, which is implicit and has remained largely unnoticed, assumes that the phenomenon of misrepresentation reduces to the phenomenon of misperception broadly conceived. These two tenets of causal-informational models are in fact closely connected. Thus, the central role played by projectable properties, rather than nomic correlations, in the production of accurate representational content is better understood as soon as one is willing to consider that at the core of the phenomenon of mental representation some basic process of re-identification is at work. Re-indentifying an item is not the same as having two identical perceptions of the same item from one encounter to the next. Rather, the ability to represent something depends on the ability to identify it as being the same under different circumstances, thanks to different percepts and this in turn is what the phenomenon of robustness embodies.

1.4 Misperception versus Misidentification

In the preceding section, the reflection on the phenomenon of robustness and the analysis of the difficulty encountered by causal/indicator theories in accounting for the identification of individuals have provided reasons to challenge the idea that an appeal to type to type nomological connections were either necessary or sufficient for establishing the distinction between true representation and misrepresentation. Causal/indicator theorists anchor the co-occurrence of recurrent signs with the entities they signal to the nomic necessity of some natural law with the hope of securing the representational link between the two, one being a reliable indicator of the other. They seem to assume that the main goal of any model of mental representation is to carefully specify a set of ideal conditions that could secure the path from what is to be represented at the source all the way to the perceptual mechanism of the organism having a representation through the mediation of properly generated signs carrying unaltered information. Such a strategy, I have argued, results from a philosophical confusion about the nature of intentional content which leads to identifying the phenomenon of misrepresentation with a phenomenon of misperception broadly construed. I do not mean to deny that perception plays a key role in the production of mental representation. However, proper perception does not guaranty against misrepresentation, while on the other hand, true representation does occur on the basis of misperception, sometimes even in a consistent and systematic manner.[7] This is so because the act of representing accurately depends on an ability to

[7] The idea that the production of true representation on the basis of misperception is not limited to the fortunate result of isolated events can be illustrated as follows. Young tennis players are commonly trained with special balls softer and more manageable than the standard ones used in competition. In order to avoid confusion, manufacturers have adopted a color code, making the training balls orange while the

properly track down, identify and re-identify entities. It does not depend on the fact that the perception of such entities occurs according to any specific and perennial causal process from one occasion to the other. This consideration helps us to understand why the failure of Fodor's ADT hypothesis in answering the challenge of misrepresentation reveals that some of the central features of causal/informational models are philosophically unsound. The phenomenon of misrepresentation does not reduce to a problem of misperception broadly conceived, even if it is true that misrepresentation often results from misperception. Many different kinds of disturbances can occur at any stage along the causal path connecting the presence of the represented item to the production of the mental token that such a presence ends up triggering. Each of these disturbances is a potential occasion for the production of misperceptions. Misperceptions can manage to alter the process of proper identification producing misrepresentation. That is the case, however, precisely because, the process of re-identification conceived as a key aspect of a proper representation does not reduce to the occurrence of a new proper perception. An operative procedure of identification needs to be already in place, quite independently of the actual causal process of perception responsible in each occasion for the production of a given mental tokening. It is only, so to speak, by hijacking such a procedure that misperception can end up generating misrepresentation.

standard ones are yellow. This policy is of no help to a friend of mine, a colorblind tennis coach. Fortunately, training balls tend to be slightly fluffier, a difference that the colorblindness of my friend makes much more vivid to him than it really is. As a result my friend is able to accurately identify orange training balls on the basis of the systematically produced misperceptions of yellowish-fluffy balls. I should point out that it is only after I had pressured him to reflect on how he was able to make such an accurate discrimination despite his handicap that my friend became aware of the actual process by which this was accomplished.

The overt line of reasoning supporting the ADT hypothesis delivers its result only by relying on an implicit and illegitimate appeal to what I should refer to as the normative element embedded in the production of mental representations. For the time being, I will limit myself to what I take to be the less controversial aspect of this normative dimension, the one which even causal theorists commonly welcome in their explanations.

Hence, the ADT hypothesis according to which the nomic connection between 'horse' and horses is the robust one—other connections being only parasitic—is justified by Fodor himself with the reason that, after all, 'horse' is *supposed* to mean HORSE. The expression 'is supposed to' marks the presence of the normative element. A highly complex network of causal paths is turned into a univocally semantic relation by cleaning the informational channels from the noise resulting from the interferences of unwelcome accidental/contingent causal connections. This strategy can only succeed if the meaning of HORSE has been already independently established, waiting to be used as an evaluative standard to which the result of the causal process of mental tokening is to be compared.

For, at the level of the intricate network of causal connections between represented items and the mental tokens they generate, it is the causal path linking the token 'horse' to the long disjunctive list of its potential referents (among which are cows in the dark looking horsey and such) that should be regarded as the robust one. It is only from the standpoint of an observer who takes the normative challenge of mental identification to be already independently settled, that some of the redundancies securing the causal path between represented items and their mental tokenings, can be read off as potential equivocations. Only then, such redundancies can be seen as disrupting, rather than supporting, the semantic relation between mental representations and the things they represent.

Fodor's ADT takes full advantage of this confusion by developing a strategy of bait and switch between semantic relation and nomological causation in providing a solution to the disjunctive problem. By not acknowledging the true nature of the normative element contained in mental representations, causal theorists are lead to let this normative dimension play an implicit role in the specification of mental content that is illegitimate.

The following analogy may help to illustrate such considerations. For reasons that will become clear in chapter four, it is important for the reader to keep in mind that, as such, thought experiments never constitute conclusive evidence. Hence, the value of the analogy below is only to make more vivid the argument just developed, not to supplement it. Imagine that a police detective is trying to identify the person whose name has been put down on a hit-list to be the next target of a renowned professional hit-man. To get this information is crucial to the detective inquiry for it is the missing piece in building a case against a dangerous gang. Assuming the detective is not concerned with saving the life of any of these criminals but only with accessing this valuable information that is the identity of the target, he could decide to wait for the shooting to happen and discover who the victim was afterward.

This strategy however depends on the reliability of the shooter's skills. In a perfect case scenario, the hit-man will be infallible, always fulfilling his contract. Under such circumstances, the detective could confidently identify the name of the supposed victim, the one on the hit-list, by relying on the identity of the actual dead body. This situation duplicates what CCT takes to be the case with normal tokenings (by contrast to wild) of mental symbols. The infallibility of the hit-man metaphorically pictures the nomological connection that, for example, in Dretske's indicator semantics, is supposed to secure the informational path from its source to its mental representation. The infallibility of the hit-man would justify the identity of the target to be reliably specified in virtue of being the actual cause of the hit-man firing.

However because it cannot be assumed that the shooter will never miss, killing an unintended target, this mode of identification remains unsatisfying. Sticking to it would make it impossible to account for any mistake on the part of the killer and will force the detective to wrongly conclude that the identity of the victim can always be traced back to the name written on the list. That is Dretske's problem of error.

In the same way, Fodor's CCT more refined account of misrepresentation as a disjunction problem rather than a mere problem of error, can be restated using our detective story by reflecting on an alternative scenario. Here, an unexpected victim is discovered by the detective, putting him in the position of having to decide whether this is the result of the hit-man misfiring or the proof that the hit-man was given two different contracts. In the first situation, the occurrence of the misfiring leading to the death of the wrong person mimics the production of a wild tokening generating a case of misrepresentation. In the second situation, the killing of the victim is the equivalent of a case of disjunctive meaning occurring as a result of two independent causal paths with no contingent causal interference occurring.

Following the logic of Fodor's ADT, the way the detective could decide which is which would be by appealing to counterfactual reasoning. Hence, suppose that the detective has reasons to think that a contract has been put on Jones' head and is surprised to find out that Smith has been killed by the hit-man. Discovering that Smith bears a very peculiar physical resemblance with Jones, the detective will be entitled to reason that if it were not for the fact that a contract was on Jones' head, Smith would have not been killed whereas the reverse is not the case. Furthermore, notice that if it is true that an independent contract was put on Smith's head, then, it is also true that Smith would have been killed whether or not he resembles Jones. Applying the ADT principle shows that the first case is clearly the equivalent of a case of misrepresentation, the second case of disjunction of meaning. The logical structure of our detective story therefore strictly duplicates the one of mental representation as conceived by causal representational models.

The value of the hit-man example comes now to the fore in helping to point out the inconsistency at the core of such models. This can be intuitively suggested by the two different ways in which the hit-man may fail to fulfill his contract: he can target the right person and accidentally miss, eventually killing someone else, or he can successfully hit the person he is aiming at while targeting the wrong person.

Causal theories don't have the resources to distinguish between these two scenarios in a consistent manner. In fact, to succeed in accounting for the phenomenon of misrepresentation along the lines of ADT, causal theorists need to reject the contrast just drawn between these two scenarios. Yet, to account for the intentional content of mental representation in a causal theory fashion, they need to implicitly endorse it. Let us make this clear by reflecting on each case.

Case A, misrepresentation resulting from misperception:

In such a scenario, the detective's guess about the name that must have been written on the list ends up being wrong because the hit-man misses his target and kills someone else. In this case, misrepresentation seems to result from misperception in a manner conforming to the causal theorist expectations. Notice that this basic plot can be further developed by offering precisions on the actual causes of the hit-man failure. The shooter may have been distracted by a bird, betrayed by his rifle or simply nervous and unfocused. A causal and/or indicator theorist will be quite right to reject these specifics as irrelevant. Once misrepresentation is conceived as the result of misperception broadly construed, it does not matter whether the information flowing from its source gets corrupted from the start, at mid-course or only when registered and processed by the observer's representational mechanisms. In short, it makes no difference whether the killing apparatus, that is the hit-man, his rifle and his ammunition, or the uncooperative conditions of the scene of the shooting, or both are to be blamed. The phenomenon of misrepresentation is robustly insensitive to the many different ways in which the misperceptions responsible for producing it may occur.

Case B, misrepresentation resulting from misidentification:

The shooter properly hits the person he was targeting. In such a scenario the detective will be wrong about the name on the list only if the shooter targeted the wrong person. The causal theorist needs to argue that this scenario, when properly construed, is yet another case of misperception. The fact that Smith has been killed instead of Jones leads the detective to wrongly assume that Smith's name was on the list. The only reason why someone could think that case B differs from case A, a causal theorist must argue, will be by taking too literally the narrative features of such a thought experiment, wrongly focusing on the hit-man's psychology.

Of course, from the shooter's point of view, missing Jones, whom he has perfectly recognized in the crowd and hitting Smith instead, amounts to being clumsy, while killing Smith because he has taken him for Jones (that is for the supposed target) amounts to being confused. Hence one may intuitively imagine cases of misperception to be different in nature from cases of misidentification. However, the causal theorist will argue that one should not be carried away by the narrative aspect of the analogy. The hit-man with his rifle is just an analog for the complex set of perceptual mechanisms needed to process the information responsible for generating mental content. It has been agreed already when reflecting on case A that whether misrepresentation results from the malfunctioning of such a complex apparatus or from uncooperative conditions of the wordy setup or both, does not make a substantive difference. The causal theorist will maintain that strictly speaking, case B is really just an elaborate and somewhat misleading version of case A.

I agree with the causal/informational theorist, that it is always possible to read case B in such a way that the difference with case A becomes irrelevant. Under such an interpretation, any of the two modes of failure respectively described by each scenario is a sufficient condition for misrepresentation, because each of them disturbed the causal path carrying the information to be processed. In addition, the causal theorist and I have agreed that this disturbance could be generated in a different way each time a misrepresentation occurs.

It is true that under such a reading any of the two modes of failures respectively described in scenarios A and B will end up in misrepresentation (or in our analogy in a failure from the part of the hit-man to honor his contract). However, despite what the causal theorist seems to believe, establishing a set of ideal conditions under which neither A or B or any similar causal disturbances could never occur will not be enough to discriminate misrepresentation from true representation.

True, in case B, the fact that Smith was killed is partially explained by the fact that Smith resembles Jones. This explanation is only partial, however: for consider the standard situation where Jones is the actual victim of the shooter, a situation analogous to the production of an accurate mental representation. The fact that the hit-man successfully got the job done is not explained by the mere fact that Jones resembles Jones, for Smith resembles Jones too. The causal informational theorist insists that contrary to Smith, Jones resembles Jones for the right reasons. But it cannot be, as the causal theorist wants it to be, that the 'right reasons' for resembling Jones consist in having the higher-property of actually being Jones. For having the property of being Jones is not the same as having the property of being the designated target. It is with regard to this last property only that whether or not the contract has been fulfilled is to be decided.

The real challenge about mental representation then, is not to explain what is needed for Jones to look like Jones in order to be properly perceived as Jones. Rather, it is to explain what it is for Jones to be identified as the designated target in order to end up being the right victim. In the case of the hit-man scenario this may not be a problem for a list of names is already an intentional item. The link between the proper name 'Jones' and its reference is a semantic one, and by having his name on the hit-list Jones is the designated target as a result of a social practice or rule observed by gang members. Notice that, being the target in that sense—that is being the expected victim, the one that the hit-man is supposed to kill—is a *normative* property *not* a *causal* one. This normative property of being the target, that is, the designated victim, does not reduce to the property of being the target in the sense of merely finding oneself at the other end of the hit-man's rifle.

The dead person is always a victim but not always the proper one and not simply, as the causal theorist would have it, because the causal connection between the person targeted by the killer and the actual victim is not always reliable or properly secured. In fact, the nature of the causal link between the victim and the shooter seems totally irrelevant for explaining *why* the hit-man fails to fulfill his contract when he does, even if it serves to explain *how* this failure occurs. For the causal connection between the name on the list and the victim can be said asymmetrically dependent on the link between the name of the list and its actual reference only if this second link can be semantically established prior to, and independently from, the causal connection between the expected target and the actual victim. The general principle behind such remarks is that causal informational theories could at best contribute to explain how misrepresentations are generated but not why they are misrepresentations.

Let us make this point clear by returning one final time to our detective story. It is only thanks to the pre-established operative procedure of writing down on the hit-list the name of the expected victims that, after reading the name 'Jones' on the list, the hit-man is able to properly identify Jones as the man he must kill that day. This is why misperceiving Smith as Jones ends up in Smith's death. The fact that Smith looks like Jones explains the misperception which itself is responsible for the occurring of the misidentification of the supposed target. That alone, however, does not explain in what consists the misidentification itself, even if it is true that without such a misperception of Smith, the misidentification would not have occurred. This is revealed by the fact that properly perceiving Jones will not lead to Jones' death if it was not for the fact that Jones has been independently identified as the designated target, a normative property he gets from having his name written on the list, not a causal property he possesses thanks to his observable physical features. The fact that such observable features are themselves intrinsic to him because of the higher-property he possesses of being Jones, and have been conveyed to the hit-man perceptual apparatus in the most reliable fashion, is beside the point.

The general attitude of causal theorists is to be skeptical about the very idea that intrinsic to the production of mental representation is a normative element. On the face of it, normative features do not square easily within a fully naturalist framework. However, rather than being a problem for naturalism, the kind of minimum normativity involved in the intentional content of mental representations directly derives from the fact that the representational mechanisms of living organisms are natural processes fulfilling natural needs. Such mechanisms do not generate representations as mere objects for inner mental contemplation on the part of the organism in which they are generated. Rather, these mechanisms produce mental representations as means for successfully coping with the external world.

Ultimately, what decides the accuracy of the content of a mental token is not its mode of production but whether it properly contributes to the success of the organism in adjusting to the situation responsible for triggering such a token, the way it should, namely whether or not it accomplishes its function. This explains why most of the contemporary models of mental representation are in fact elaborate models best described as causal-functionalist models. One could therefore reasonably assume that the normative dimension that has been shown to be lacking in the bare causal informational explanation is to be accounted for by the functionalist analysis that commonly complements causal informational models.

The next chapter will be entirely devoted to the examination of the standard notion of function involved in functionalism in order to find out whether or not such a notion is in fact adequate for coping with the normative aspect of mental representation.

CHAPTER 2: FUNCTIONALISM AND ITS LIMITS

In the light of the problems faced by the causal informational treatment of mental representations, notably the problem of error, it is necessary to examine the extent to which functionalism, the other dominant perspective, provides a more satisfying type of model when offered as a complement to or as a substitute for bare causal informational theories. Chapter two is therefore dedicated to functionalism, with a close analysis of the standard notion of function at work in functionalist models of intentional content.

This chapter is divided in four sections. Section 2.1, *Representations and Malfunctions in Functional Role Theories*, defines functional role semantics and explains how a purely functionalist approach undermines the notion of representation. The so-called two-factor theory is then introduced as an attempt to salvage representationalism in the context of a semi-representational model combining causal and functional elements. Aside from the technical problems attached to each of them, the different versions of functional role semantics depend on a common functionalist notion of function. The remaining part of section 2.1 presents some reasons for suspecting that such a notion of function may fail to provide a sound principled distinction between well-functioning and malfunctioning devices. Section 2.2, *Functionalist Kinds and the Problem of Membership*, argues that the standard functionalist notion of function, as such, does not have the resources to properly define the extension of any given functional kind of entities or the criteria for membership to such kind. Functionalism overcomes such problems only by helping itself with an intentional reading of the notion of function which, in the context of a naturalist account of mental content, is illegitimate. This last point is argued for in section 2.3, *Functionalism without Functions*, on the basis of a careful comparison between the traditional theoretical definitions of natural kinds of entities and the functionalist definitions of functional kinds of devices. The main result of this analysis is twofold. First, without the appeal to some normative evaluation external to the functionalist explanation itself, functionalism cannot properly account for the phenomenon of malfunction. Secondly, and more surprisingly, successful functionalist explanations do not rely on the actual fulfillment of any function of its own on the part of the device under study. Section 2.4, *"Functioning as" Functions versus Proper Functions,* applies these results to Cumming function-analytical type of explanation, a landmark example of functionalist analysis, and contrasts this type of explanations with teleological explanations and more precisely with Millikan's notion of "proper function".

2.1 Representations and Malfunctions in Functional Role Theories

In the first section of chapter one, I explained how and why functionalism constitutes another dominant approach to mental representation. The functionalist approach is sometimes offered as an alternative, sometimes as a complement to causal-informational theories. Functional role semantics (sometimes also called conceptual role semantics), that is functionalism applied to mental representation, is the idea that mental contents, and not merely mental states, are to be identified by their functions. Functionalism itself divides into two basic kinds: causal functionalism and computational functionalism. Under causal functionalism, it is the causal role of a given mental representation that determines its content, under computational functionalism, its computational role. This distinction applies to functional-role theories of mental representation (Cummins, 1995, pp. 114-125).

The functional role of a given thought T represents the disposition of this thought to interact with other thoughts, sensory states, and motor skills, which all contribute to the overall economy of the organism's behavior. A great number of causes and effects are likely to be involved in the production of T or to result from such a production. Many of these causes and effects are non-semantic. Block has suggested as an example of such effects, the fact that happy thoughts may bolster up one's immune system, promoting good health. The conceptual role of T is what remains of the functional role of T once non-semantic causes and effects have been discounted.[8]

In the preceding chapter, I have argued that the problem of error couldn't be solved by bare causal-informational theories because to misrepresent is to fail to represent what *should be* represented, it is not to fail to comply with what is supposed to constitute the ideal, the standard or even the more frequent set of conditions and means for generating reliable representations. The idea that there is more to accurate mental representation than true perception properly produced or information reliably channeled and processed, could be seen as making the case for the adoption of functional role semantics (Harman, 1987). By contrast with causal-informational theories, conceptual role theories identify the content of a given thought in terms of its functional role, that is,

[8] In addition to the traditional justifications for adopting functionalism in general, the adoption of conceptual role semantic is generally justified by the fact that many terms in human language seems definable not individually in respect to the way they connect to the world, but only by reference both to other terms and to their particular location in the net of representations they belong to. Conceptual role theorists are keen on supporting their view by directing our attention to theoretical terms such as the ones of 'atom', 'spin' or 'quark' which seems to acquire their meaning only in connection with one another and by the role they play within the conceptual framework of some scientific theory.

in term of what such a thought accomplishes in taking part in the general economy of the thinking activity of the organism. For it is a disconcerting feature of a bare causal/informational account of intentional content that it ignores what it is that a mental representation does for the organism that has it or what it is that the organism that has it does with such a representation. On the other hand, a concern for such questions is central to conceptual role semantics. According to such a view, how a mental token interacts with other mental tokens within the general economy of the mental life of the organism and helps to modify its behavior is precisely what determines the semantic content of that particular token as a member of a given functional type.

That being said, a purely functionalist account of mental states seems to make the representational dimension of mental content ultimately irrelevant. This can be best seen in the case of computational functionalist semantics. Under this version of functional-role semantics, the content of a given mental state is determined by the computational role that such a state plays in the implementation of cognitive processes, which are themselves understood in non intentional terms as abstract functions carried out on the basis of syntactic rules partially responsible for producing the organism's behavior. In such a context, it is difficult to understand in what sense a reference to the representational aspect of mental content is needed at all in accounting for the functional role of any mental state. As Robert Cummins remarks:

> Functionalism thus encourages us to think of mental states as causal mediators, and hence as things whose explanatory role is to make possible a story about the causation of behavior. When a functionalist comes to

> think of representation, therefore, it is inevitable that the problem should be posed thus: What is the role of representation in the causal mediation of behavior? Once we combine this picture with computationalism, it is hard to see how the concept of representation could do any serious work. (Cummins, 1995, p. 124) [9]

If the contents of mental tokens are entirely defined by the computational function resulting from their syntactic properties, then the semantic properties of mental representations are merely supervening properties, which are doomed to be causally idle. It is not even clear what makes them representations and justifies that they would be distinguished from the many other non-representative side effects resulting from the computations of the syntactic engine operating in our heads.

Hence, a purely functionalist account of mental content may lead to the adoption of anti-representationalist views. Anti-representationalism seems a consistent view to hold for an eliminativist. As explained earlier, eliminativism regards folk psychology as hopeless. Most of the terms in our ordinary vocabulary such as 'thought', 'belief', 'will', 'attention' and the like are at best totally ill-conceived or worse, do not refer at all. In section 1.1, I presented the reasons why this view has been resisted by most naturalists. To its credit, though, eliminativism is perfectly consistent with full-blood physicalist

[9] As suggested by Cummins himself, Stich's original version of computationalism model (1983) can be seen as a perspective that fully acknowledges the anti-representationalist leanings of functionalist explanations, consistently carrying them to their ultimate consequences.

naturalism. In fact, eliminativism does not so much offer a theoretical model for mental representations as it gives a theoretical justification for why no such model is needed.

By contrast to computational-functionalist theories, causal theories criticized in the precedent chapter, seem to accommodate intentional representationalism much more convincingly. According to causal theorists, notions such as the ones of 'beliefs' and 'desires' do refer and their references have a key explanatory role to play in the success of the behaviors of the organism in which they occur. This representationalist understanding of thoughts carries on into language and verbal behavior. Typically, causal theorists will start from the default position that the meaning of a word is given mainly, although not quite entirely, by its referent and the meaning of a sentence, again, mainly although not quite entirely, by its truth-conditions.

I have myself explicitly endorsed representationalism by assuming from the start that intentional content was best approached in terms of mental representation. What could be the point of having thoughts if they were not of a representational nature and what could those thoughts possibly mean if not what they represent?

Furthermore, the rejection of representationalism flies in the face of Darwinian evolution. Thus, as Michael Devitt and Kim Sterelny argue, the best reason for endorsing representationalism comes from the fact that:

> No other approach offers a ghost of an explanation of why humans and other animals think and talk at all. Humans, and to a lesser extent other animals, have invested a great expense in a large brain. We use significant chunk of that brain to make and listen to noises. Why have such

> capacities evolved? The only explanation available is a representational one. In Godfrey-Smith's felicitous phrase, accurately representing one's world, as it is and as it could be, is a "fuel for success" (1996: 172). Creatures whose representations map the world are advantaged in satisfying their needs. This explanation is not without problems. But it is without rivals. In particular, no narrow or internalist theory of thought can explain why the capacity for thinking is adaptative, for that explanation must exhibit systematic relations between creatures so endowed and their environments. But narrow theories look only inside the mind. In sum, representationalism solves an overarching evolutionary problem. (Devitt & Sterelny, 1999, p. 208)

Hence, a functionalist willing to salvage the representational aspect of mental representation could reason that the causal-informational theorist defends a largely externalist account of meaning that makes it possible to fully embrace representationalism. On the other hand, bare causal-informational theories tend to disregard the ways mental representations are used by the organism, how they help the organism to adjust its behavior in order to satisfy its needs; something that, the functionalist believes, is constitutive of their meaning. As the result the causal theorist so to speak puts the theoretical constraint in the wrong place, namely on what constitutes the proper conditions of production of mental tokenings rather than on what constitutes a successful implementation of their functional role.

A model known as two-factor conceptual role semantics (Block, 1987; Field, 1977), based on a of division of labor's principle, is supposed to offer a view with the right kind of moderate externalism so as to take advantage of the best features of both functionalist and causal-informational models. Some modified version of Putnam's famous Twin-Earth thought experiment will help to see how such a model is supposed to work.

Thus, while water is (H_2O) on earth, it is (XYZ) on Twin-Earth. Being abducted during his sleep, Oscar, an average terrestrial being, will wake up on Twin-Earth not noticing any changes. Oscar will keep thinking of water and use the term 'water' in speeches as he did before but the truth-value of at least some of his thoughts and sentences will have changed.

As it is often the case with thought experiments, the conclusion to be drawn is uncompromisingly ambiguous. Putnam wanted to conclude that meaning was not in the head. Conceptual-role theorists concluded instead that a commitment to strict externalism makes meaning irrelevant in accounting for speech dispositions and intentional behavior. They inferred that it was the 'narrow content' of thoughts and linguistic items—a constitutive part of the psychology of the organism not an objective property of the external world—which was causally efficient and had explanatory force in accounting for the organism's actions. Narrow content was therefore responsible for the true meaning of WATER.

If narrow content were really all that there was to meaning, than conceptual-role semantics would be uncompromisingly non-representational; an unfortunate feature that may well out-weigh any of the advantages of functionalism. But according to two-factor conceptual role semantics, meaning is constituted of two elements: an internal component which corresponds to the narrow content of thoughts and an external component which corresponds to its referential/truth-value. The functional role of the thought is supposed to account for the internal component of meaning, its causal connection to the world for the external component.

Sadly, the two-factor theory faces many problems on its own, the most obvious one being the difficulty to produce a satisfying elucidation of the nature of the connections between the two factors. It also shares some common problems with standard conceptual role theories. For example standard conceptual role theories and two-factor theories alike seem to be forced into the alternative of arguing for the soundness of the analytic/synthetic distinction with respect to meaning or conceding semantic holism.[10] These questions are still in debate.

That be as it may, the important point for our present discussion is to decide whether or not one has reasons to think that the functionalist approach supporting conceptual role semantics could be more successful in solving the problem of error than causal-informational approaches. Functionalism understands mental representations as the result of mechanisms implementing a given function that is itself cashed out in terms of the *powers* or *dispositions*, actual or potential, of the entity that possesses such mechanisms. When applied to intentional content, this standard functionalist account of the notion of function, I will now argue, renders the problem of error impossible to solve for, to put it crudely, no system can ever fail having the functions it has. No function of

[10] On this problem see chapter three of *Holism: A Shoppers' Guide* (Fodor & LePore, 1992).

the cognitive apparatus of the organism can ever fail to accomplish its job, in our case the production of the proper mental content, while remaining at the same time the very function it was when this apparatus was working properly.

This problem is the functionalist equivalent of Dretske's problem of error in causal-informational semantics. As earlier, this problem can be seen as a particular aspect of Fodor's disjunctive problem. Hence faced with two mechanisms, A and B, displaying similar but somewhat divergent sets of behaviors, functionalism does not have the resources to decide whether, A and B are functionally equivalent, one of them failing to properly achieve their common function, or whether, A and B are simply endowed with two distinct functions that each of them fulfills properly.

This problem remains unnoticed as long as one only focuses on the mechanisms of artifacts which have been designed to serve a specific function. It seems easy enough to tell whether or not a computer or an air conditioner is properly functioning because it is easy to tell what each of these devices is supposed to do. However, while artifacts result from the deliberate creation of a designer who had their expected uses in mind all along, biological mechanisms, representational mechanisms included, do not. Contrary to artificial devices, living organisms do not come with instruction manuals attached. It is not simply that, as a result, the actual function of natural devices may be more difficult to identify than that of the artificial devices. It is also that some principled distinctions, which are required for the mal-functioning versus well-functioning contrast to be operative, are explicitly and independently given in the case of artifacts, while they are not in the case of biological mechanisms.

Artificial devices generally constitute self-contained entities entirely devoted to well-specified tasks with a limited degree of flexibility in their behavior. Hence, an air conditioner unit is supposed to maintain the temperature of the room within a certain level and when it successfully does so, it does it each time in an identical manner, by going through the same physical processes. In addition the AC unit has been conceived so that its active interaction with its surrounding is strictly limited to what is needed for performing its task. Its design protects it also against any other possible interactions that are not helping or could disturb such a task. Finally, what constitutes a proper surrounding for the air conditioner is straightforwardly derived from considering the use for which it is designed. A small portable AC unit is fit for cooling the bedroom of a family home, not the warehouse of a big factory. Owners of artifacts are responsible for making sure that the proper conditions of use match the requirements specified by the manufacturing company. Typically, user's manuals contain a clause discharging the manufacture of any legal responsibility in case of malfunction resulting from inappropriate conditions of use.

The above considerations explain why the boundaries between, on the one hand, the set of processes corresponding to the implementation of the artifact's function and, on the other hand, the set of proper conditions in the surroundings which permit such processes to run smoothly can be clearly specified in the case of artificial devices. Hence a failure on the part of the device can be clearly distinguish from a failure resulting from inappropriate conditions of use. Such a principled distinction proves to be much more elusive when standard functionalism is applied to biological devices. Artifacts can be subjected to a functionalist analysis on the basis of the causal powers and disposition of the type of device under study and from the point of view of the expected benefit for the owner of this artificial device. The fact that the evaluation is user-relative does not constitute a problem, for saying that the artifact was conceived with its functional use in mind is the same as saying that it was conceived with the user's benefit in mind.

By contrast, the actual purpose of a biological device and the normal conditions under which this purpose is supposed to be fulfilled cannot be established by reference to the standard functionalist notion of function alone. A normative reading of what constitutes the expected purpose of such a device is therefore added to the functionalist explanation, along with a set of idealized conditions under which the fulfillment of such a purpose is supposed to be guaranteed. Without the addition of such a normative expectation, the functionalist explanation would be empty or circular. With the help of this normative reading, the functionalist explanation becomes substantive but is not fully naturalistic anymore. Such an explanation depends on an intentional reading of the notion of purpose which cannot be accounted for in standard functionalist terms.

Because they both rely on the standard functionalist notion of function, one may suspect that, the one-factor as well as the two-factor approaches to intentional content fail to provide a sound principled way for discriminating between misrepresentations and true or proper representations. Functionalists pick and choose among representations after they have been produced, selecting the ones which agree with the expectations generated by their models. However this cannot be done by reference to the necessary and sufficient set of properties specifying the given function responsible for the production of such representations, for this definition itself is based on the entire set of the actual and potential representational outcomes, correct ones and misleading ones alike.

Thus, functional role semantics may not fare better than causal theories in providing a naturalist account of intentional content, while presenting the unwelcome tendency to support semantic holism and anti-representationalism. Yet, there is something right about the idea of appealing to the failing of a function in order to solve the challenge of misrepresentation. However, the following section will provide arguments for thinking that such a challenge cannot be successfully met on the basis of the standard functionalist notion of function.

2.2 Functionalist Kinds and the Problem of Membership

There exists, I take it, a reasonably clear distinction between, on the one hand, non-normative claims about what is the case, what was the case or what will be the case, and on the other hand, normative claims, that is, claims about what is supposed to be or ought to be the case. This distinction is clear if not in practice at least in principle. Sometimes the issue is muddled by the fact that some claims, such as predictions, are stated using a wording that makes them sound like normative claims although they are not. For example, someone looking at a grey sky in the morning may be saying "it should rain in the afternoon". But predictions are claims about facts not norms. It is just that the facts to which they are referring are facts about the future rather than present facts or facts about the past.

I have argued that there is a normative dimension to mental representation which causal-functionalist models fail to properly acknowledge. As a result, such models cannot solve the so-called problem of error and more generally cannot account for the phenomenon of misrepresentation in a fully naturalist manner. As explained in the preceding section, the best that a standard (non-teleological) functionalist can do to distinguish a well-functioning device from a mal-functioning one—hence in the case of a representational device, to distinguish the production of an accurate representation from the production of a misrepresentation—is to compare his expectation about how the device should behave with how the device is actually behaving. But "should" here cannot be read normatively; it is the "should" of prediction. From a standard functionalist point of view, what should happen is what one predicts will happen which itself is defined by what has been observed as happening on average or most of the time with the kind of mechanisms to which belongs the particular device under study. This approach is problematic for several reasons discussed below.

Standard functionalism demands that the performance of a given representational device be evaluated in relation to the function of the particular kind of devices of which it is a member. This presupposes the existence of two distinct and well-specified procedures: one by which a given kind of functional devices is established or circumscribed, another by which a particular device is identified as belonging to such a kind. It is tempting to imagine the two tasks completing one another, the second smoothly unfolding from the first. Yet, a closer inspection shows that, on the contrary, in the context of a standard functionalist analysis, these two tasks run against each other. Hence, a functionalist will want to define a kind as a set or class of devices actually fulfilling a common function due to a similar display of powers and/or dispositions. While such a procedure may help in achieving the first task at hand, namely establishing the existence of a given functional kind of devices, it renders inoperative the second task of deciding whether or not a particular device belongs to such a kind. It does so, either by ruling out in advance malfunctioning devices or by rendering their acceptance as members of the class puzzling and in need for a justification that a strictly functionalist definition of the notion of kind cannot provide.

Hence, a given kind is understood as a class C of items behaving, or at least with the disposition to behave, in accordance with some specified function F. In such a context any device (d) unable to perform F under suitable conditions must be denied its membership to C, which in the case of representational devices leads us back to the problem of error. On the other hand if (d) is accepted as a member of C in order to make room for malfunctions, then it must be acknowledged that the standard functionalist account of function needs to be both completed and amended in a way that requires justification. For, with respect to the standard functionalist perspective, the recourse to some external criterion providing a principled distinction between cases in which devices like (d) do not qualify as members of C and cases in which devices like (d) qualify as defective members of C presents all the attributes of an *ad hoc* hypothesis. No such criterion can be logically deduced from the functionalist account of the mechanism itself. Yet such a criterion cannot be stated or properly established for its own sake either, but must rather be carefully adjusted each time, that is, in relation to each new class of entities. Furthermore, solving the problem of membership from one class of functional items does not help explaining nor predicting what will happen when moving to another class, because it does not help the standard functionalist to establish a rule for what constitutes a proper criterion in general. Finally the validity of the criterion adopted cannot be tested against any counterexample for it is this criterion itself and the specific way it is applied that decide, with respect to any given class, what counts or not as a counterexample in the first place.

Hence, without the support of an external criterion stating the norm, whether such criterion is established on the basis of some social conventions, linguistic rules, paradigmatic cases, or any alternative candidates, the standard functionalist account remains incomplete. With the introduction of such a criterion the functionalist explanation tends to help itself with an implicit normative clause, treading on an intentional reading of the notion of function that violates the constraints of a fully naturalist explanation.

The question of what constitutes the boundaries of a given kind of items as defined by their common function, when such a function itself is accounted for in standard functionalist terms, represents more than a technical difficulty for functionalists. It is the symptom of the inadequacy of their philosophical perspective. It is tempting for functionalist theorists to regard the above criticisms as trading on a misleading interpretation of their view, one that pictures their definition of function as a mere stipulation, making it more arbitrary and much less substantive than it really is. Such a reaction may be comforted by the belief that a more accurate analysis will show that functionalist explanations, properly understood, are in fact composed of two distinctive elements: a theoretical definition for a given kind of items that fixes the actual extension for a given class of entities and a procedure for establishing whether or not any given item qualifies for membership to such a class. Functionalist theorists may be thinking that when properly articulated, these two elements make immaterial most of the supposed difficulties pointed out above.

Thus, the functionalist philosopher would argue that the categorization of a specific domain of entity that a standard functionalist definition establishes, does not amount to a mere class specified by stipulation but rather denotes a kind of entity actually sharing common properties which objectively command their unification. Hence, functionalists may believe that the questions attached to the meaning of functionalist definitions of kinds of devices, the actual extension of such functional kinds with respect to such definitions, or yet the criteria for membership to such kinds, are no more problematic than are the questions about the actual extension of natural kinds of entities or the proper criteria for membership to such natural kinds. If it could be shown that functionalist definitions of kinds of devices are indeed analogous to theoretical definitions of natural kinds like gold or water, then functionalists would be right in downplaying the seriousness of the concerns offered above. It is therefore necessary to have a closer look at what is involved in the definition of natural kinds in order to see to what extent this applies to functionalist definitions of kinds of devices as well.

An identification of a given kind of entity is provided thanks to a core of common properties shared by all its members. Such properties are intrinsic properties essential for membership, such as the atomic number (79) for gold or the (H_2O) chemical formula for water. One should not be confused by the fact that the procedure offered for identifying such essential properties implies the identification of some dispositional properties which help testing for membership to such kinds. In most cases, the discovery of the intrinsic properties of natural kinds on the one hand and the criteria for membership to such kinds on the other hand, are likely to be indistinguishable for any practical purpose, yet they remain distinct in principle. Kinds properly identified must not be reduced to the description of the common dispositions of their members. These two elements are indeed irreducible to one another, despite the fact that, in general, both may appear in the theoretical definitions of natural kinds. Such a distinction is reflected in part in the fact that a linguistic community is often in control of the use of some natural kind term, applying such a term consistently and successfully, long before being in command of any accurate scientific description of the necessary properties responsible for the overt dispositions of the entities belonging to this kind.

The underlying assumption here is that, in the ideal world of an achieved science, the expert's theoretical definition and the set of necessary properties this account helps to capture will eventually end up being extensionally identical. Hence, one can see how such a treatment of natural kind terms seems to provide the philosophical ground for an understanding of such terms as referring to objective kinds of entities and how theoretical definitions of such natural kinds of entities have a substantive meaning rather than a mere conventional one.

Functionalist could argue that a similar approach applies to functional kind terms which in a like manner refer to functional kinds of devices and conclude that the status of functionalist definitions of a kind of device is no more problematic and no more arbitrary than the one of theoretical definitions of natural kinds of entities. What is wrong, then, with such an analogy? Providing a detailed answer to this question will constitute the main topic of the next section.

2.3 Functionalism without Functions

First a certain ambiguity exists between two possible understandings of the expression "functionalist explanation" that needs to be clarified. Thus, the expression "functionalist explanation" sometimes names an explanation of the nature of a certain kind of item or the working of a given kind of device in terms of the fulfillment of a specific function (sense1, hereafter). On other occasions, the expression refers to a certain procedure that functions as a test for establishing membership to its kind for a given item or device (sense2, hereafter). It is easy to get confused here, especially because each one of these two readings refers to one of the two aspects constitutive of any functionalist explanation.

Notice that a similar ambiguity exists also in the context of the theoretical definition of natural kind terms. Hence in the ideal context of a perfectly achieved science the expert's theoretical definition of a given natural kind (K) will at once perfectly capture the essential properties making up (K) and will offer a full-proof procedure by which any given item (k) could be properly classified as a member of (K). In the case of natural kinds, like gold or water, which are as close as something can be to having real essences, the distinction between what makes an item the kind of thing that it is and what sort of test helps establishing the membership of this item to a given kind almost collapses entirely. With such cases, some standardized procedure decides for a given item (k) its membership to the kind (K) by testing for some dispositional properties of (k) directly resulting from (k)'s intrinsic constitutive features following a (sense2) reading of the functionalist explanation. These intrinsic features, like the atomic number (79) for gold, are also constitutive of the theoretical definition of (K) itself following a (sense1) reading, making the distinction between (sense1) and (sense2) largely irrelevant for any practical purpose. Yet, the distinction is real and persists even within the restricted context just discussed, where it is rendered almost invisible.

In the context of functionalism, the constant shifting between (sense1) and (sense2) readings of the account of functional kinds of devices makes it difficult to notice the recourse to an external normative criterion. This is particularly true with models of mental representations which identify misrepresentation with the failure on the part of representational devices as defined in standard functionalist terms.

It therefore will prove useful to exercise our discriminatory skills within the restricted context of the theoretical definition of physical or chemical kinds of entities first, before trying to clarify similar confusions in the much more complex context of the functionalist account of the representational devices of biological organisms. The well-known story of Archimedes and the golden crown of king Hiero will serve our purpose here. Inspired by the spilling of water as he is submerging his body in the bathtub, Archimedes imagines a test by which could be settled the question of whether or not the crown ordered by the king is really made of pure gold. Archimedes' test provides an operational definition of the natural kind gold based on the specific density of this metal, even if only specified relatively, that is by comparison with the density of some other proxy metal, silver in the present case. As the story goes, the crown fails the test, revealing the dishonesty of the goldsmith. But of course, nothing can ever fail to be what it is and, in that respect, the expression "failing the test" applies to the crown only metaphorically.

In fact, theoretical definitions of natural kinds like gold or water do not rely on the fulfillment of any function on the part of the item tested for membership to these kinds.

This is hardly surprising since it makes little sense to talk about a sample of metal trying to be made of gold or trying to behave as gold, let alone succeeding or failing in doing so. At this point, a natural reaction is to move away from a (sense1) reading of the theoretical definition and to turn to a (sense2) reading of the failure of the crown to pass the test. Under this alternative interpretation, it is the test that failed when applied to the crown, not the crown that failed when put to the test. On reflection however, this second reading proves to be no more satisfying and no less metaphorical than the first. Of course it is always possible for a test to fail in a very literal sense and the reference to the fool-proof test of the perfect expert envisaged earlier in relation to natural kinds is no more than an idealization.

Yet, in the case of the king's crown, there is no good justification for assuming that the test did not succeed. Archimedes reasoning about the relationship between the volume of liquid displaced by the emerged crown and the respective density of the two metals, gold and silver, is sound. In addition, Archimedes' technical skills combined with the fact that a man's life was at stake provide sufficient reasons for assuming that the procedure was carefully conducted as well. Furthermore, as noticed earlier, it is only metaphorically that the crown can be said to have failed the test. This metaphorical way of talking points, nonetheless, to the very real fact that the crown was not made of pure gold, something revealed only because Archimedes' test successfully fulfilled its function. The test, or more accurately its result, is a failure only in the sense that the king's expectation to acquire a true golden crown was not met, that is, only because the goldsmith failed to behave as he should have by serving his king honestly. Such considerations about the king's expectations or the violation of the goldsmith's moral duty, introduce a normative evaluation that is not part of any theoretical or operational definition of gold.

I do not mean to suggest that Archimedes' solution, or for that matter, the contemporary theoretical definition of gold as having the atomic number (79), need revision. I take both to be perfectly adequate. Defining gold as having the atomic number (79) and water as having the chemical formula (H_2O) are theoretically sound definitions of natural kinds under a (sense1) reading of the theoretical definition. Under a (sense2) reading of the same definition these are effective procedures for deciding in favor or against membership to such kinds. But precisely because nothing in the world can ever fail to be gold or water but simply is or is not a sample of gold or a drop of water, the failure when it occurs is always on the side of the theorist's expectations, never on the side of the object or phenomenon that the theory is supposed to explain.

If it is granted that the theoretical account of natural kinds is correct, then the notion of failure makes sense only in relation to some external normative expectations on the part of the theorist or more generally in relation to some social rules, practical needs, particular interests or culturally shared representations. This fact is rather easy to notice when the failure revealed by the test occurs in the context of Archimedes' story.

The introduction of such a normative evaluation still occurs, but remains much more difficult to detect, in the context of a functionalist account of a function attached to a given kind of devices. Referring to the failure of a given device to carry out a certain task, or to properly play its part in a complex process, sounds less contrived than talking of a silver crown failing to behave as a golden crown and, more generally, is less problematic than describing a given item as failing to be something that it is not.

This sense of reassurance is deceptive. For as long as the notion of function is understood in standard functionalist terms, the success of the functionalist explanations about the functional role of a given kind of devices no more involves the fulfilling of any function of their own on the part of the members of this kind than it does with samples of gold or drops of water. Similarly, the occurrence of the notion of failure or error, assuming that it is not referring to a problem in the theory itself, is no less dependent on the normative expectations of the theorist in the context of the functionalist explanations of a device's function than it is with King Hiero's crown. It is just that, here, the *ad hoc* introduction of an external normative element of evaluation is made much more elusive. The constant shifting between (sense1) and (sense2) readings of the functionalist explanation is itself made more difficult to track when the explanation is intended to capture the behavior of a device by reference to what is presented as its function rather than the identification of an item by reference to the necessary properties presented as constitutive of its nature. The following comparison will help to make this clear.

In the first situation of the theoretical account of a given natural kind and the identification of its members, we notice the following elements at play:

> [1] The notion of 'a class of items' is understood in terms of membership in a common kind. Kinds are themselves defined thanks to a theoretical definition making reference to a set of necessary properties, in conformity to a (sense1) reading of the theoretical definition.

> [2] A procedure is spelled out for identifying an item as being or not being a member of a given class in conformity to a (sense2) reading of the theoretical definition.

> [3] An item can be tested for membership by being subjected to the procedure defined in [2]. This item can be said to fail the test only metaphorically and only under the assumption that the test itself, as defined in [2], succeeded.

By comparison, the situation is much more complex in the case of the functionalist definition of a functional kind of devices:

> [1]' The notion of 'a class of functioning devices' cannot straightforwardly be defined by reference to a set of necessary properties shared by all the devices carrying out such a function. This is resulting from the fact that functions are characteristically capable of multiple-realization.

In section 1.1, we explained how and why it was largely this feature of multiple-realizability that made functionalist models of mental representations appealing to a vast majority of theorists. A first consequence for our present discussion is that a functionalist explanation cannot refer to a set of necessary properties common to a given kind of items in order to delimit the actual domain of reference for a given function. In that sense, there is no direct equivalent to the (sense1) reading of the theoretical definition of kinds of entities in functionalist explanations of functions.

The difference between the theoretical definitions of natural kinds of entities and the functionalist definition of functional kinds of devices runs deep. This results from the fact that the phenomenon of multiple-realization does not occur only with respect to the physical constitution of the different devices carrying the same function. It concerns also the actual mechanisms or causal process by which such a function is implemented. In short, the very same function can be realized by different kinds of processes which can themselves be implemented on the basis of different kinds of physical supports. A functionalist definition of a given kind of function cannot be established on the basis of some set of necessary properties that would be shared by all the devices carrying out such a function. The functionalist explanation must rather refer to a given way of functioning, that is, it should spell out for a given kind of function F a theoretical definition of "functioning as" an F.

> [2]' The procedure established for identifying a device as being a member of its class in conformity to (sense2) cannot be simply that the device tested is actually functioning as predicted. For if that were the case, malfunctioning devices would be denied membership along with devices belonging to alternative functional classes. First the device tested needs to be identified as a member of a class of devices that have generally demonstrated something like an ability and/or a disposition to behave in

> accordance with the theoretical definition of “functioning as” introduced in [1]’.

Such a reference to abilities and/or dispositions is problematic for several reasons. The notion of ability (or capability) is both extremely complex and ambiguous in itself. An organism may have the causal power to function as an F but not the proper means. It may have the proper means but not the know-how; the know-how but not the opportunity; and so on. How many and which of these different elements are required for an organism to be recognized as having the ability to function as an F?

Furthermore, notice that even under the assumption that a list of such necessary conditions is eventually established, and given the context of an ideal situation where all these necessary conditions do obtain, so that there is no doubt remaining about a particular organism’s ability, this organism may still have no disposition whatsoever to put such an ability to use.

The reverse is also true, at least under a certain understanding of the notion of disposition. For one may demonstrate a constant disposition to act in a given way without ever succeeding in doing so, due to a total lack of ability. Of course, the notion of disposition itself can always be redefined so as to imply the ability to succeed under the right conditions. In this context, (sense2) of the functionalist explanation will refer to the spelling out of a procedure for testing the functioning of a given device (f) understood as the ability for (f) to function as an F, through the testing of (f)’s dispositions to behave as expected under the right circumstances. Nevertheless, it remains that having both the ability and the disposition to play a particular role in a complex process, to fulfill a given task or to carry on a certain job, even under the right conditions, is not the same as succeeding in doing so, unless the notion of what constitutes “the right conditions” is rendered entirely vacuous or circular by being strictly identified with situations of actual success.

> [3]’ A device (f) which, when tested, does not function as an F, counts as a mal-functioning device, rather than as a device with an alternative function or no function at all, if and only if (f) has been already identified as a member of the class F of devices. Membership to such a class being itself established in terms of a “functioning as” definition, (f)’s membership to F implies that (f) needs to have already demonstrated the ability and disposition to function as an F at least on some previous occasions.

As observed above, a device (f) may have both the ability and the disposition to function as an F and nonetheless never behave in accordance with the functionalist’s expectations. Conversely, a consequence of the functionalist approach seems to be that a device (g) behaving in a way that satisfies the theoretical definition of functioning as an F, even if only on a few occasions and out of sheer luck, would be granted *ipso facto* membership to the F class. Without the introduction of some normative evaluation external to the functionalist explanation itself, functionalists are condemned to treat false positive cases as genuine ones, while at the same time denying membership to their own class to some genuine—although sometimes defective—devices. Under the functionalist definition of what constitutes a class of functional devices, the distinction between mimicking and actually implementing a given function F seems to disappear and with it the principled distinction between borderline cases and deficient ones.

2.4 "Functioning As" Functions versus Proper Functions

The main result of the comparison between a theoretical account of the nature of a given kind of item and a standard functionalist account of the functioning of a given kind of device can be summed up as follows. In the first situation, in conformity with a (sense1) reading, a theoretical definition of kinds of entities specifies what constitutes the extension of a given class of devices in terms of the necessary properties for membership. Such properties are actual properties of the items belonging to such a kind. In that respect, (sense1), that is the theoretical definition, is independent from (sense2), the procedure for testing membership. In the case of natural kinds, the theoretical definition of, for example, water as being (H_2O) captures what Saul Kripke (1980) would like to describe as an *a posteriori* necessary truth about water. The point is that the definition is substantive but does not involve any function, while the procedure used to test an item for membership involves the particular function of the test itself rather than any function of the part of the item tested.

In the second situation, for reasons explained above, (sense1) needs to make reference to the notion of "functioning as" an F, an expression which remains epistemologically ambiguous for it is not clear whether such a theoretical definition refers to an independent function F, actually owned by devices which are belonging to the F class, or whether the definition establishes which set of dispositions will grant a device

(f) the status of member of the F class. This ambiguity carries on to (sense2) to the extent that the procedure for testing membership focuses exclusively on the functioning of (f) understood in terms of (f)'s capacities and dispositions. It remains uncertain whether the behavioral patterns exhibited by (f) when they meet the theorist's expectations are better understood as what actually constitutes (f)'s membership to the class F, or if they rather serve as reliable signs of the fact that (f) does in fact possess F as its function. The interesting thing to notice here is not merely that it is difficult to tell which interpretation is correct, but that deciding in favor of one interpretation or the other has no impact on the functionalist explanation itself. Even when dealing with devices which do have a function on their own, that is, independently of any intentional reading on the part of the theorist, functionalist explanations make no appeal to such a function in order to account for the functioning of such devices.

Let me illustrate this last point by reflecting on Cummins' description of his own version of functionalist explanation as found in *Functional Analysis* (1975). Cummins explains that:

> If something functions as a pump in a system *s* or if the function of something in a system *s* is to pump, then it must be capable of pumping in *s*. Thus, function-ascribing statements imply disposition statements; to attribute a function to something is, in part, to attribute a disposition to it. If the function of *x* in *s* is to ϕ, then *x* has a disposition to ϕ in *s*. For instance, if the function of the contractile vacuole in fresh-water protozoans is to eliminate excess water from the organism, then there must

> be circumstances under which the contractile vacuole would actually manifest a disposition to eliminate excess water from the protozoan that incorporates it. To attribute a disposition *d* to an object *a* is to assert that the behavior of *a* is subject to (exhibits or would exhibit) a certain regularity: to say *a* has *d* is to say that *a* would manifest *d* (shatter, dissolve) were any of a certain range of events to occur (*a* is put in water, *a* is struck sharply). The regularity associated with a disposition—call it the *dispositional regularity*—is a regularity that is special to the behavior of a certain kind of object and obtains in virtue of some special fact(s) about that kind of object. Not everything is water-soluble: such things behave in a special way in virtue of certain (structural) features special to water-soluble things. Thus it is that dispositions require explanation: if *x* has *d*, then *x* is subject to a regularity in behavior special to things having *d*, and such a fact needs to be explained. (Cummins, 1975, pp. 757-758)

Cummins' remarks make it clear that his functional analysis treats in a uniform manner the action of the contractile vacuole in fresh-water protozoan and the disposition of a particular object *a* to shatter when struck sharply or to dissolve when put in water. Hence his functionalist explanation covers indiscriminately cases which, from a teleological point of view to be defined soon, must be carefully differentiated. Thus, a teleologist like Millikan will identify the action of the contractile vacuole eliminating the excess of water from the organism as an example of a device that is fulfilling its purpose by successfully carrying out its proper biological function. By contrast the dispositional behavior of object *a* to shatter when struck sharply or to dissolve when put in water does not involve the fulfillment of any function of its own. In that respect, Cummins' opening statement according to which, "if something functions as a pump in a system *s* or if the function of something in a system *s* is to pump, then it must be capable of pumping in *s*", deserves particular comments.

Notice that for a particular device *x*, functioning as a pump in a system *s* is not necessarily the same as having the function to pump in *s*. The second expression implies that *x* is endowed with a function on its own whereas the first expression has no such implication. Under a functionalist theoretical definition of the notion of "functioning as" it seems perfectly justified that the property of functioning as a pump be granted to *x* if and only if *x* is capable of pumping in *s* as required by Cummins-like analysis. (This comes as the direct consequence of the analysis developed in [3]' of section 2.3.) But the same is not necessarily true when *x* is described as having the function of pumping in *s*. Millikan's teleological perspective, the nature of which will be the subject of an extensive analysis in chapter three, helps to render this difference particularly clear. This explains why Millikan, whose teleological account of proper function (or purpose) has been sometimes awkwardly confused with the so called "function analytical" type of Cummins' explanations[11], promptly reacts to Cummins' statement by noticing that:

> It is of the essence of purposes and intentions that they are not always fulfilled. The fact that we appeal to purposes and intentions when

[11] An example of such confusion can be found in Beth Preston's *Why is Wing Like a Spoon? A Pluralist Theory of Function* (1998).

> applying the term "function" results directly in ascriptions of functions to things that are not in fact capable of performing those functions; they neither function as nor have dispositions to function as anything in particular. For example, the function of a certain defective item may be to open cans; that is why it is called a can opener. Yet it may not function *as* a can opener; it may be that it won't open a can no matter how you force it. Similarly, a diseased heart may not be capable of pumping, of functioning *as* a pump, although it is clearly its function, its biological purpose, *to* pump and a mating display may fail to attract a mate although it is called a "mating display" because its biological purpose is to attract a mate. (Millikan, 1989, pp. 294-95)

Millikan's notion of "purpose" is supported by her elaborated teleological theory of proper functions, which will be discussed in sections 3.3 and 3.4 of the next chapter. The notion of a proper function is complex but one simple aspect of this notion is particularly relevant to the present discussion and concerns the use of the adjective "proper", in the often misconstrued expression "proper function". A function is qualified as "proper" by Millikan not because it is properly carried out or because it has been properly theoretically defined. Rather, a function is qualified as "proper" because it is a function which is proper to the functional device under study; a function that it is the device's purpose to fulfill quite independently from any theorist's expectations. This is not a "functioning as" description ascribed to the device in the context of some functionalist models.

The ocean functions as a thermostat by cooling the air in July and warming it in November, making Arcachon, a small location in France by the Atlantic Ocean, a very pleasant place to stay. Such a function, however, is not a proper function of the ocean. By the same token, the glass in the window of my room may be fragile, but being fragile is not its function. Dispositions can be the subject of functionalist descriptions leading to statements of the form if (f) is a member of the functional kind F, then under standard conditions (f) should behave as an F. Again, the 'should' here is descriptive and/or predictive, not normative. It is only by imposing some intentional reading of the notion of function that standard functionalist theorists are able to turn this 'should' into a normative one. At this point, the normative expectation that this 'should' is referring to does not concern the proper function of the device under study anymore. This is why when a piece of glass does not break according to functionalist expectations, or when the temperature of Arcachon drops way below average in Fall, no failure on the part of the glass or the ocean is to be pointed out. The failure is on the theorist's side and the error when there is one is in his model or data. But things are entirely different when the contractile vacuole fails to eliminate excess water from the organism the way it should. In this case, 'should' has a normative value and the failure is really a failure on the part of the device to fulfill its own purpose independently of any theorist's expectations or predictions.

This is not to say that standard functionalist accounts of the functioning of a given class of items are necessarily inadequate or useless and should always be replaced by teleological ones. It is quite the opposite. For example, the type of function-analytical explanations offered by Cummins often provides illuminating descriptions of the functioning of elaborate artifacts, computing systems, investments patterns or ecosystems' cycles. Thanks to such function-analytical explanations, accurate predictions can be made and new operative tools and strategies conceived. Substituting a teleological approach to the standard functionalist one when dealing with such issues would most of the time be simply impossible or clearly inadequate. The problem is rather that the success of the standard functionalist approach depends, in each case, on the introduction of an additional normative element thanks to which the theorist can provide the adequate interpretation of the function under study. Such a normative element is reflecting the theorist's interest or purpose. Without the background of expectations entailed by the adoption of such an intentional reading on the part of the theorist, functionalist explanations would not have the theoretical resources to adequately delimitate the extension of a given class of functional items; to distinguish between borderline devices and deficient ones; to discriminate between mere simulations and actual implementations of a given function; to establish the actual abilities of a given device based on its behavioral dispositions, and finally to distinguish situations in which functions are ascribed to devices in order to explain and predict their patterns of behaviors from situations in which it is the actual theory-independent function of such devices that helps explaining such behavioral patterns.

In the case of mental representation, following Millikan's model of teleosemantics, I will argue in the coming chapters that such a normative element needs to be traced back to the purpose that representational devices actually possess as the result of their evolutionary history. Instead, functionalism provides an account of such devices by ascribing to them a-historical functions defined in terms of powers and dispositions on the basis of counterfactual situations. In causal-functionalist models of mental representations, these counterfactual situations are carefully chosen to compensate for the fact that, after overlooking the normative dimension attached to the phenomenon of mental representation, functionalists are forced to implicitly appeal to some external semantic criteria of demarcation between true representations and misrepresentations, in order to make their model work. In light of the above analysis, one understands why the challenge of misrepresentation faced by causal-functionalist models of mental content is not a technical problem that some refinements in future versions in the already quite long history of functionalism could eventually put to rest.

CHAPTER 3: THE TELEOSEMANTIC PERSPECTIVE

Standard functionalist notions of function, I argued in the last section of chapter two, cannot handle the problem of misrepresentation because they are incapable of capturing the normative aspect of intentional content adequately, that is in a way that agrees with naturalism. A teleological notion of functions, I will argue now, is needed. The present chapter focuses mainly on such functions as they are conceived and developed in Ruth Millikan's model of teleosemantics. Why Millikan?

Ruth Millikan's model occupies a predominant place within the field of teleosemantics. Where many teleosemanticists only gesture toward some reasonable way to generate a naturalist account of mental representations, Millikan provides a detailed explanation from her own comprehensive perspective, the foundation of which was rigorously established in her seminal work *Language, Thought, and Other Biological Categories* (LTOBC, hereafter) in 1984, and which has been continuously developed and deepened since.

Her detailed treatment, however, comes at the cost of introducing an entirely new technical apparatus of elaborate notions that resist any superficial reading. Millikan's work is extremely demanding. Almost each page of LTOBC introduces a concept original and difficult to grasp. The role of each original concept can be fully understood only after one has gained an adequate overview of the entire model. This imposes a constant back and forth reading that makes it easy to lose track of the argument. A second difficulty has to do with the many areas of philosophy that are covered. The reader is expected to possess more than a decent understanding of intricate issues in semantics, logic and metaphysics, as well as a reasonable grasp of experimental psychology, the study of animal behaviors and the current state of affairs in evolutionary biology. Last but not least, is the difficulty resulting from the complexity and originality of Millikan's thinking. She offers a radical perspective that cuts across most of the well-established dichotomies between different schools of thought, challenging many of the received categories of traditional analytic philosophy. In light of all these considerations, one can fully understand why when writing about the teleosemantic program, proponents and opponents alike tend to focus on more manageable versions that are easier to master, summarize and discuss.

In addition, while central to teleosemantics, Millikan's model cannot be described as mainstream either. It is generally considered bold and radical by many who share her naturalist perspective. Godfrey-Smith (1996) expresses a fairly common view among them when he writes, "Ruth Millikan's theory is one of the most immodest, as it uses the same apparatus to explain why inner states of beetles can be about mates, and why the English word "mate" means mate. It is also fairly immodest to try to explain all mental representations (including human beliefs) at once, leaving public language and the like for another theory." (p. 176).

Hence, Millikan's model is often perceived as too ambitious in its scope and too uncompromising in its means. In contrast with Millikan, many theorists are indeed tempted to narrow the scope of teleosemantical explanations or to allow non-strictly teleosemantical features to play a role in their own models. At least this is how their different moves appear when observed from a Millikanian perspective. The fact that Millikan's views are referred to as "High Church" by Karen Neander (1995) or "pure" teleosemantics by Pierre Jacob (1997) reveals the uncompromised nature of her project in the eyes of other teleosemanticists. (Of course, presenting the positions of other teleosemantists as mere departures from Millikan's model would be clearly inaccurate, for most of them represent original and independent creations.)

Yet, for all its inherent complications, I am convinced that Millikan's highly elaborate and uncompromising project should operate as the center of gravity of this research. Because of its fully developed nature, the model addresses many of the problems that critics of teleosemantics have raised. By contrast, other teleosemanticists are generally concerned with one particular set of issues or another and remain brief or silent with respect to the rest.

This chapter is divided in five sections. Sections 3.1 and 3.2 focus on explaining the true nature of teleosemantics with the intent to discard most of the common misconceptions attached to this philosophical perspective. Section 3.1, *The True Nature of Teleosemantics,* explains that the only thing that theorists really need to share to qualify as teleologists is an understanding of the phenomenon of misrepresentation as resulting from the failure of some representational device to fulfill its teleological function. Section 3.2, *The Price of Teleology*, analyzes under which conditions a teleological perspective could be added to a preexisting model of representational content and to what extent such an addition is likely to impact the model. Dretske's model of informational indicator semantics is used here as a case study.

Sections 3.3 and 3.4 provide a detailed analysis of Millikan's teleological functions. Section 3.3, *Millikanian Functions*, analyses the notion of "proper function" which is at the heart of Millikan's model, while section 3.4, *Derived Functions and the*

Novelty Issue, analyses how the complex apparatus of Millikan's teleological functions manages to account for the diversity and creativity of mental representations on the basis of historically-based, and therefore fairly conservative, types of explanations. Finally, section 3.5, *From Natural Signs to Intentional Representations*, introduces the distinction between intentional and non-intentional signs and establishes the role that the consumer perspective and the set of Normal conditions for the production of intentional signs respectively play in determining the nature and content of intentional representations.

3.1 The True Nature of Teleosemantics

As noticed in the introduction, a great number of misunderstandings and inaccurate renderings of teleosemantics are found in the literature, comforting some ill-conceived criticisms against it. It is tempting to picture teleosemantics, the view that purports to naturalize mental content on the basis of the etiological functions of living organisms as shaped by Darwinian evolution, as an alternative to causal and functionalist models. Such a presentation, however, if not carefully amended, will be clearly misleading.

The first important thing to realize is that, strictly speaking, the reason why different teleological models can rightly be classified as teleological *is not* because they share some common account of the way mental representations are produced. In that sense, the situation differs greatly from what is the case for causal-informational models and functionalists models. Rather, what is shared by all teleological theories is a common treatment of the phenomenon of misrepresentation and more generally a common strategy to answer the challenges offered by Brentano's notion of intentionality.

“Intentionality” is a Middle-Ages term from the Scholastics that was reintroduced by Brentano (1874) in *Psychology from an Empirical Standpoint*. The exact meaning of the term in Brentano’s writings has been an everlasting subject of controversy among historians of philosophy.[12] Aside from the intricate discussions among specialists of this issue, and more directly relevant to the present discussion, is the fact that Brentano’s notion of intentionality has also developed a life on its own. As it often happens in philosophy with fruitful concepts, such a notion eventually established itself as an intellectual short-cut that encapsulates in one simple expression a complex perspective. Hence, Brentano’s notion of intentionality is often used as a convenient starting point in the debate over mental representation.

So what exactly constitutes the shared understanding of Brentano’s notion of “intentionality” as referred to in contemporary writings? This notion needs to be broken down into two main elements: aboutness and the capacity to refer to non- existent objects. Aboutness characterizes intentional states in the sense that a mental state is intentional to the extent that it is a representation of something, or that its content refers, accurately or not, to a particular object, “object” being understood here in the most inclusive sense. Among such objects is postulated the existence of a peculiar sub-set that is said to deserve closer attention, the one made of non-existent entities. For the capacity to refer to non-existent objects is claimed to be the second distinctive feature of intentionality. Together these two features, aboutness and the ability to refer to non-

[12] To get a sense of such controversies see notably *The Philosophy of Brentano* (McAlister, 1976).

existent objects, make up intentionality. Furthermore, intentionality so understood is offered as the distinctive mark of the mental.

Beside the question of 'aboutness' itself, Brentano's account of intentionality presents the naturalist philosophers with two apparently distinct challenges. Theorists need to explain both how it is possible for mental representations to be false and how it is possible for mental representations to be empty, that is, to be representations of non-existent objects. While teleologists may, and in fact sometimes do, largely disagree on what constitutes the best theory of mental content, they present a united front in addressing this double challenge. They claim that misrepresentations, whether they are inaccurate representations or empty ones, result from the failure of the mechanism responsible for generating these representations to fulfill its purpose, that is, to accomplish the function it has been designed for or selected for.

This explanation, Millikan argues, applies to any theoretical account of mental representation that one is willing to defend in the first place. Hence, describing the theoretical contribution of Tilly, the teleologist, to the problem of mental content, Millikan writes:

> What a teleological theory of content does is to take some more basic theory of content, point out that the application of that theory to actual creatures requires idealizing them in certain ways, and then offer the teleological principle to explain which idealization is the right one to use in interpreting intentional contents, namely, the one that fits how the cognitive systems were designed or selected for operating. You give your

> naturalistic analysis of what a true or correct representation is like, and Tilly merely adds that systems designed to produce true representations don't always work as designed, claiming that correctness in perception and cognition is defined by reference to design rather than actual disposition. (Millikan, 2000, pp. 229-230)

Several remarks are in order here. Firstly, Tilly claims that misrepresentations must be distinguished from true representations by reference to the design of the representational system responsible for producing them. This is a substantive claim only if Tilly can offer a comprehensive account of what it is in general for a biological mechanism to be designed for producing inner representations but also, in each situation, what it is that the representational mechanism under study has been designed for representing. It is only after she has offered a positive account of what constitutes the correct answers to those questions, that the teleologist, it seems to me, rather than being just a theorist subscribing to the teleological approach, becomes a teleosemanticist with a positive model of mental representation in her hands.

However, addressing such questions requires a prior theoretical account of the kind of functions for which it can be said that a biological mechanism, feature or item, has been selected for fulfilling. For this reason any teleological model of mental content presupposes the existence of the equivalent of what in Millikan's model is referred to as 'proper functions'. Most of this chapter will be devoted to the detailed study of Millikan's account of proper functions.

Secondly, the proper function of some mechanism (or trait) defines what this mechanism has been selected for doing, what it is supposed to do, its purpose. In that limited sense the notion of 'proper function' is normative. At the end of the previous chapter, I concluded that functionalist notions of functions could not successfully handle the problem of misrepresentation. The basic reason for this short-coming, I explained, was that functionalism hopes to establish the distinction between true representation and misrepresentation, in a counterfactual manner, by reference to some idealized set of dispositions or causal powers. These idealized conditions are themselves inferred from the observation of the actual powers and dispositions commonly displayed (most of the time) with high frequency by (most of) the members of a certain type of organisms. As the result, the functionalist account of misrepresentation is circular and ultimately vacuous under a certain reading, substantive but not fully naturalist under another.

By contrast with the functionalist notion of function, the teleological notion of proper function is not causal or dispositional. To use a landmark example, the proper function of the heart is to pump blood rather than, let us say, to make a regular noise, despite the fact that the heart has both the disposition and the causal power to make a regular noise no less than it has to pump blood. Pumping blood is the proper function of the heart; this is what the heart is supposed to do, because it is what it has been selected for by evolutionary forces. The fact that their hearts were pumping blood efficiently explains the survival of the ancestors of the actual heart-owners, hence the existence of such an organ in current members of the species. The heart example helps also to see that, in addition of being modestly normative, the notion of proper function is essentially historical. It is only by reference to the history of its ancestors that it makes sense to talk about the proper function of some feature, mechanism, trait or behavior displayed by an organism. The historical dimension of the notion of ‘proper function’ and its consequences on the nature and content of mental representations will be discussed in chapter four.

The notion of ‘proper function’ must not be understood in terms of statistical average either. As Millikan often mentions, the proper function of sperm cells is to reach the ova, despite of how few actually succeed in doing so. According to Millikan the same teleological analysis applies to animals’ visual systems or belief-forming mechanisms. Looking at the proper function of a given device helps to establish the purpose of such a device, what it is that this device does when working properly. It is by reference to such a normative expectation that biological mechanisms can be said to be malfunctioning or, in the particular case of representational mechanisms, to misrepresent.

Finally, the notion of ‘selection’ needs to be refined if it is to be properly used in teleology. Following Elliot Sober’s analysis, one should distinguish being merely ‘selected’ from being ’selected for’. To illustrate the distinction, Sober (1984) imagines a vertical cylinder containing several horizontal dividing partitions from top to bottom. Each partition has holes of a given size, the holes getting smaller the closer the partition is to the bottom. The cylinder is used as a sorting device for balls of different sizes, the biggest balls getting stuck at the top and only the smallest ones reaching the bottom. As it turns out, balls of different sizes are also of different colors, the smallest ones being green. Hence by selecting the smallest balls, Sober’s cylinder also selects the green balls. However, while color is merely selected, size is selected for. It is the size of the balls, not their particular color, that explains why they end up at the bottom of the cylinder, and therefore it is what they have been selected for. When referring to the teleological function of a given mechanism or trait, teleologists refer to the function that has been historically 'selected for' by the process of natural selection in that sense.

Hence, teleologists identify the occurrences of misrepresentation with situations where representational mechanisms fail to achieve the purpose supported by their teleological function. In that respect, and contrary to what Brentano's account of intentionality tends to suggest, empty representations do not require special considerations, for they are not representations of some mysterious non-existent objects, they are just radical failures to represent. These empty representations are misrepresentations nonetheless, that is, they remain representations, although totally deficient ones, and the mechanism producing them remains a representational mechanism as well.

The philosophical insight behind such a position is not as counterintuitive as it appears, for it is not unlike the reason why when an electronic failure prevents an ATM machine from providing you with the cash you need, its screen is more likely to read something like "OUT OF ORDER" than "NOT AN ATM MACHINE". Notice by contrast that someone willing to provide a purely functionalist account of the heart's function in standard functionalist terms of power and dispositions will be forced to conclude that, strictly speaking, nobody could ever die from a heart attack.

The link between purpose and norm remains mostly underestimated by opponents of teleosemantics. The same theorists also often overtly reject the essential connection established by teleologists between these notions of purpose and norm on the one hand, and the historical nature of proper functions on the other hand. I do not mean to imply that such a resistance to the idea of connecting norm with history is a direct consequence of the tendency to under appreciate the importance of the link between purpose and norm. As I will discuss in chapter four, causal theorists have, or at least believe themselves to have, strong independent reasons for objecting to the central role that teleologists wish to give to history in accounting for mental content.

Millikan (1995, p. 23) has suggested that language itself may contribute in part to obscure the essential connection between function and purpose and more decisively between these two notions and norms. This has to do with the fact that in ordinary language, purposive behaviors are generally referred to by employing 'success verbs' rather than 'trying verbs'. For example, there is no verb to describe the action of the hit-man of our first chapter when he is shooting at someone but his rifle gets jammed and does not fire. Trying to shoot does not count as shooting, hence the need to preface the verb 'shooting' by the expression 'trying to' to convey the exact idea of what happened. Linguistic terms employed to refer to behaviors tend to categorize them by reference to their effects rather than by reference to their purposes and when a reference to their purposes serves as a basis for classification, the terminology tends to focus on fulfilled purposes rather than tentative ones. As a result, our ordinary linguistic practices tend to obscure the connection between purpose and norm.

Furthermore, the fact that most success-verbs don't have trying-verbs counterparts, the way 'believing' is the trying counterpart of 'knowing', contributes to making the teleological treatment of mental representations counter-intuitive in yet another way. Verbs of perception, lacking real trying-verbs counterparts, end up being used equivocally such as to cover both succeeding and merely attempting behaviors.

Millikan (2004, p. 65) gives the example of expressions like 'hearing voices' or 'seeing pink elephants'. Without such an equivocation in the use of perception-verbs, the temptation to postulate that a mental object has to exist for one to hear it or to see it, when there is nobody speaking at the moment, nor any pink elephant in the room, could be better resisted. Once both the lack of many try-verbs and the equivocal use of some success verbs in describing successful and unsuccessful behaviors alike have been pointed out, the idea that, for example 'seeing' a pink elephant when drunk amounts to successfully representing some mysterious mental entity, rather than merely failing to properly represent anything at all, loses its force. Brentano's treatment of intentionality seems to have mislead him to conclude that explaining 'aboutness' was the flip side of accounting for the existence of special entities before thought that were not part of the external world and that the capacity of the mind to connect directly with such entities was the distinctive mark of the mental. By accounting for the phenomenon of misrepresentation in term of the failure of some representational mechanism to fulfill its teleological function, teleosemantics get rid of such mysterious mental entities and the need to explain their role in the formation of intentional representations.

3.2 The Price of Teleology

To avoid common confusions about teleosemantics, one should bear in mind that the teleological theorists' account of mental representation does not reduce to the teleological aspect of their models. Conversely, the defense of teleology must not be misconstrued as a commitment on the part of teleologists to some common model of mental representation. As a result, while one should expect teleologists to agree on the fact that misrepresentation occurs each time a representational mechanism fails to fulfill its purpose, one should not expect teleologists to agree on what it is actually that this mechanism must effectuate in order to properly represent something. This is so, partly because different theorists may have different ways to construe the notion of 'proper function' (or what the equivalent teleological notion turns out to be in their respective models) and partly because they may have different accounts of how this function is normally carried out.

In the light of these considerations, one better understands why Millikan insists that teleosemantic models are riding piggyback on basic and possibly divergent models of intentional content. However, it seems to me that this suggestive picture of teleology could be misleading too if not cautiously qualified. One should not be made to believe that a teleological perspective could be easily added on top of any model of representation in order to provide a better treatment of misrepresentation. Furthermore, one should not imagine that such an addition will leave the basic representational model untouched. Teleology does not come for free. While it is true that teleology, as such, does not compete with any positive account of mental representation nor adjudicate between them, teleology nonetheless imposes some severe constraints on these models, and to some extent alters their original nature.

Central to the teleological approach is the idea that the production of mental representations needs to benefit the organism if not always, at least in some occasions. This is the rationale behind the very existence of the representational mechanisms historically carried over generations of organisms. Let us just reflect on some important ways in which the clause that representations be beneficial ends up impacting models of mental content; such an impact being sometimes even greater than what some of the proponents of these models who are willing to endorse teleology are ready to acknowledge.

In chapter one, we have seen that causal models of content were mainly focusing on explaining how mental representations were generated but were surprisingly oblivious of how these representations were related to the organism's needs. Adding a teleological component to their models will force causal-informational theorists to address more directly the question of how the production of representations serves the organism, what it is that such representations command, direct, suggest, and help the organism to do or restrain it from doing. The adoption of a teleological approach to misrepresentation will not merely help theorists to deepen their analysis of the role of intentional content. It will also alter the nature of their explanation of mental representations.

One more time, we can usefully resort to Dretske's indicator semantics to help illustrating this point. As a result of his own critical analysis of the problem of error, Dretske knows that purely informational models of indicator-semantics do not have the theoretical resources for providing a decent treatment of misrepresentation. His complete model, therefore, evolves into a teleological-informational version of indicator semantics. Now, and aside from any teleological import, Dretske's informational account of indicator semantics is supposed to provide a basic account of the content of mental representations. As we explained, teleology as such cannot be substituted for any positive model of mental content. Dretske's informational explanation of what gives its content to representational items depends on Dretske's account of natural signs as indicators. An indicator indicates (that is reliably carries information about) a state of affairs, not merely when the occurrence of this indicator and the presence of the state of affairs coincides, but when the connection between their respective types is nomically supported by some natural law or, at least, is subjected to a statistical frequency that makes the probability of the presence of one in the presence of the other equal to 1. As expected, Dretske (1995) justifies the addition of teleology to his indicator semantic model by referring to the challenge of misrepresentation. The connection between the two is pictured by him as follows:

> Representation is here being understood to combine teleological with information-theoretic ideas. If the concept of representation is to do a useful job in cognitive science, if it is to be used, in particular, to illuminate the nature of thought and experience, it must be rich enough to allow for misrepresentation. It must include the power to get things wrong, the power to say that something is so when it is not so. This is what the teleology, the idea that something having an information-carrying function, is doing in the present theory. It captures the normative element inherent in the idea of representation. Since an object can retain a function even when it fails to perform it...a device can retain its indicator function...even when it fails to provide the information it is its job to provide. There is information without functions, but there is no representation without functions. (Dretske, 1995, p. 4)

Dretske’s own description of the role of the teleological element in his model already shows that teleology does more than just handling the problem of error. The teleological analysis permits the identification of the indicator’s “information carrying function” and Dretske goes as far as claiming that without such a reference to the indicator’s function all there is is information, not representation.

Yet, one wonders whether such a reading of the role of teleosemantics may not ultimately question the soundness of the central notion of information offered by indicator semantics as the basis for mental content. The nature of the connection between types of states of affairs and types of indicators is supposed to be an objective feature of the world. Information as defined by such a notion will not be sensitive to the function it may help to implement in relation to a particular organism. Notice that, a model of bare indicator semantics will describe an indicator as carrying the exact same information that Dretske’s teleological model describes this indicator as carrying.

Hence, Dretske’s point is certainly not that what the consumer does with the received information determines what such information is about. Rather, Dretske is saying that without a reference to the fact that it is the teleological function of such an indicator to carry information, this indicator will not be an indicator but just a reliable sign of the presence of a certain state of affair. But then, it is not clear how an appeal to the teleological function of the indicator could change its status and turn it into a genuine representational item, when this teleological function is described by Dretske as an “information carrying function”. Carrying information in the reliable way specified by Dretske’s notion of information makes a given signal an indicator, but being an indicator in that sense, is something the signal is, not something the signal does. It is therefore rather misleading to refer to the "information carrying function" of signals for it is hard to see how such signals could ever be endowed with the function of being what they are. Maybe what Dretske's has in mind is not the proper function of the indicator itself but the proper function of the producing-mechanism responsible for generating such an indicator.

According to the benefit clause embedded in the teleological perspective, if the producing-mechanism of indicators has been selected for, that is, if it is truly the teleological function of this mechanism to generate indicators, then it has to be that the production of indicators benefits the organism (or at least has benefited the ancestors of the organisms) some of the time. It has to be that producing representations makes a difference that merely carrying information does not. It follows that, contrary to what Dretske seems to assume here, the distinction between mere information and full-fledged representations cannot be drawn by simply pointing at the function of the producing mechanism of indicators. An appeal to what indicators, as representational items, actually help to effect is required. Dretske is aware of the problem but, it seems to me, is divided between his commitment to his indicator semantics model and the modifications that a teleological approach seems to impose on the notion of information supporting his theory. The internal tension between Dretske's indicator semantics model of mental representation and Dretske's teleological treatment of misrepresentation can be detected in the following:

> Ruth Millikan...has stressed the point that representations are dependent on *consumers*. I agree. If there is nothing to use (consume) the information provided, nothing can acquire the function of providing it.

> Nonetheless, the way information is consumed does not determine representational content. It does not tell us what the representation is a representation *of*. That is determined by *what* information the system acquires the function of providing. (Dretske, 1995, p. 187)

In the context of this quote 'the system' Dretske refers to is the 'information delivery system', that is the producing system of signals, along maybe with the signals it produces. The claim that the producing system has been selected for producing signals sounds reasonable enough, providing that something is said to explain how signals themselves, benefit the organism. Yet, to derive from this that the content of a representation is determined by "what the system acquires the function of providing'' is problematic, given Dretske's account of what is an indicator. This last claim seems to imply that it could be the teleological function of the producing system to make sure that the correlation between the type of signals (ambiguously dubbed indicator by Dretske) that benefit the organism and the states of affair they represent does hold. To have such a capacity, the producing mechanism would need to function in a manner similar to the way a performative utterance is said to function according to Austinian speech-act analysis. The producing mechanism will need to have the capacity to increase the chances for a given state of affairs (S) to be the case by producing indicators that pictures (S) as being the case, the way financial experts can sometimes manage to influence the market to move in one direction by delivering a public statement about what their analyses show the trend to be. I take it that it is a similar concern that Millikan (1995) has in mind when she notices that "a problem with Dretske's view is that it is hard to see how it could be the function of any biological device literally to *effect* the production of one of his “indicators”” (p. 129).

Our previous analysis of causal-informational models, in particular Dretske’s indicator-semantic model, has shown that the notion of information offered in such models is at the same time too strong and too weak. It is too strong because the constraints put on the way types of indicators and types of states of affairs need to be connected for the indicator to be representational rule out the use of many natural signs in the production of representations. We have seen before (section 1.3) how such excessive requirements were making, for example, the identification of individuals problematic. It is also too weak, because it is not clear that simply looking at the information reliably carried by an indicator in conformity with such requirements will be enough to determine its representational content. For example, it seems that it will also be necessary for the information to be properly shaped (or coded) so as to be read by the organism and used accordingly. Only information properly formed informs properly. In that respect, some of the conditions required for an indicator to play its representational role and benefit the organism depend on the nature of the mechanisms consuming, rather than producing, the information. We can see how by adding a teleological dimension to his model in order to account for the problem of error, Dretske has troubles to explain how it could be a device’s teleological function to carry information on the basis of the notion of information that constitutes the foundation of his basic account of mental representation.

It is not my intent to pursue this issue any further. I hope that enough has been said to convince the reader that endorsing teleology has a price. The main goal of these two first sections was to provide a better understanding of the true nature of the teleological perspective. A proper appreciation of teleology, I have claimed, demands bearing in mind that teleology as such does not constitute an alternative model of mental representation but also that it is not a perspective that could be added on the top of basic representational models without having the effect of reshaping them, sometimes drastically.

3.3 Millikanian Functions

The teleological answer to the challenge of misrepresentation depends entirely on the assumption that representational devices, along with many other biological devices, have teleological functions that they sometimes fail to fulfill as they ought to, in regard to what is their purpose, what they have been selected for. In Millikan's terminology, such teleological functions are 'proper functions'.

Providing a fully naturalist account of proper functions, or some equivalent notion of teleofunctions, is the most basic task of teleology. Millikan has devoted the entirety of the two first chapters of LTOBC to this task.

Many of the objections that have been raised against Millikan's teleosemantics are directed at her notion of proper functions. Most of these objections, although not all of them, have been disqualified by Millikan on the ground that, each time, the critics failed to understand the nature of the phenomenon that the notion of 'proper function' was forged to capture. In several occasions, since the publication of LTOBC, Millikan has tried to facilitate a better understanding of her position. On some occasions, she tried to provide a less detailed but more accessible account of the notion of proper functions. On other occasions, she has carefully contrasted her notion of functions with alternative notions with which they should not be confused (like, for example, Cummins-style functions).

Yet Millikan appears justified in thinking that, overall, Millikanian functions have been inadequately pictured in the literature. A list of three major types of confusions can be established here to serve as red flags to the reader as many misleading interpretations to stay away from. Firstly, despite Millikan's explicit statements to the contrary, some writers have treated the Millikanian theory of functions as a contribution to conceptual analysis. That is, they have regarded Millikan's definition as an effort to clarify the set of necessary and sufficient conditions for the proper use of the linguistic term 'function' and have mistakenly evaluated her contribution accordingly. Secondly, some theorists wrongly imagined that Millikan was mainly concerned with providing a unified notion of function that could cover the various types of functionalist explanations found in the work of contemporary biologists. Finally, yet other theorists have largely underestimated the explanatory power of Millikan's theory of functions by overlooking, misinterpreting or confusing, the intricate Millikanian notions of 'relational' 'adapted' and 'derived' functions. The central goal of this chapter is to provide an analysis of Millikanian functions as clear and complete as possible in order to remove as much confusion as possible.

Let us start by reflecting on what a complete definition of proper function requires. We already know that having a proper function is a question of history, not powers or dispositions, and also that an item possesses a proper function only to the extent that it is a member of an historical set of items that have benefitted from this function and have transmitted it over generations. The first requirement is therefore to specify the precise conditions under which a given item can be said to be a member of such an historical set of items or what in Millikan's terminology is called a "reproductively established family" (REF hereafter) of items. This, in turn, calls for a specification of the modes of production owing to which an item belongs to the same REF as its ancestors, and for that reason is similar to them in some crucial ways. Such modes of productions are named by Millikan "reproductions." An individual B is a "reproduction" of an individual A if and only if the three following conditions are satisfied:

> (1) *B* has some determinate properties, p_1, p_2, p_3, etc., in common with *A* and (2) below is satisfied. (Millikan, 1984, p. 19)
>
> (2) That *A* and *B* have the properties p_1, p_2, p_3, etc., in common can be explained by a natural law or laws operative in situ, which laws satisfy (3) below. (Millikan, 1984, p. 20)
>
> (3) For each property p_1, p_2, p_3, etc., the laws in situ that explain why *B* is like *A* in respect to *p* are laws that correlate a specifiable range of determinates under a determinable under which *p* falls, such that whatever determinate characterizes *A* must also characterize *B*, *the direction of causality being straight from A to B*. (Millikan, 1984, p. 20)

Each time one of these three conditions is stated and before the next one is introduced, Millikan provides some additional comments that need to be briefly reported here. Concerning the notion of ‘determinate property’ introduced with condition (1), she explains that a property is “determinate” relative to some “determinable” property under which both this property and a set of contrary properties fall, as in the case of *red* and its contrary properties *green, yellow* and the like, which all are determinate properties relative to the determinable property *colored*. The notion of “operative law in situ” introduced with condition (2) refers to a special law that one derives from universal natural laws by adding a reference to the actual conditions surrounding the production of a given item, B in the present case. Finally condition (3) should be understood as capturing the conditions supporting the counterfactual statement according to which if A had been different with respect to its determinate character p within a specifiable range of variation, then B would have differed accordingly as a result.

Additional details and further complications are introduced by Millikan in relation to the phenomenon of reproduction. We will refer to some of them later on when needed. However, to avoid common confusions, it is important that the role of the bare notion of reproduction be clearly identified in itself and aside from all the additional specifications attached to it. The basic role of the notion of reproduction is to explain why two different items A and B end up sharing a common property p, by explaining how, given that A has the property p, B has the property p too. In other words, using Millikan’s color example, reproduction explains why A being red B is red too. Notice that the notion of reproduction is not intended to explain why it is that A (or by extension why B) is red rather than, say, green or blue. Properties p_1, p_2, p_3, etc., that B ends up sharing with A, as the result of reproduction, are named “reproductively established properties”, and A is called a “model” for B.

I believe that a useful way to look at the notion of ‘reproduction’ is to see it as the result of Millikan’s effort to capture the general principle behind the many different patterns of reproduction and transmission of properties from one generation of items to the next, which, each in its own way, will make it the case that A and B be members of the same REF. Most of the complications I have mentioned earlier concern these many different ways by which an item can serve as a model for another, over and above the straightforward case of direct copying exemplified by the delivering of a photocopy by a copy machine fed with the original. Viruses, mass produced objects, ritualized human behaviors or linguistic items are all cases of reproduction, each according to a specific pattern that sometimes wildly diverges from the basic copying pattern. It is the value of Millikan’s notion of reproduction to bring all these types of items together under a category that rightly point out why they all qualify as members of a distinctive REF, while acknowledging the great diversity of patterns involved in their respective production.

Millikan needed to distinguish two different kinds of REF: first order and higher-order established families. The reason for the introduction of this distinction can be illustrated by returning to our heart example. All human hearts belong to the same REF, yet no human heart is the direct reproduction of another human heart from the previous generation of ancestors. Rather a certain set of genes is responsible for the production of hearts. Each token of those genes belongs to the same REF and is a copy of the same type. For that reason, while genes belong to a first order REF, hearts belong to a higher-order REF. Millikan defines first-order REF as follows:

> Any set of entities having the same or similar reproductively established characters derived by repetitive reproductions from the same character of the same model or models form a *first-order reproductively established family*. (Millikan, 1984, p. 23)

The heart example has shown the need to complete the definition of REF by adding to the first-order REF so defined, a definition for higher-order REF that reads as follows:

> Any set of similar items produced by members of the same reproductively established family, when it is a direct proper function of the family to produce such items and these are all produced in accordance with Normal explanations, form a *higher-order reproductively established family*. (Millikan, 1984, p. 24)[13]

[13] In fact Millikan offers two distinct definitions for higher-order REF. A second definition is added to the one presented here in order to account for learned behaviors resulting from training or trial-and-error procedures. Such behaviors are members of higher-order REF when it is a proper function of the mechanisms producing them to reiterate the same behavior after it has been praised or rewarded. This second definition reads as follows: "Any set of similar items produced by the same device, when it was one of the proper functions of this device to make later items *match* earlier items, and these items are alike in accordance with a Normal explanation for performance of this function, form a *higher-order reproductively established family*." (Millikan, 1984, p. 24)

Such a definition may appear problematic for it makes use of the notion of "proper function", a notion that REF was precisely supposed to help defining. It also makes use of the notion of "Normal explanation" which has not been defined yet. As long as an analysis of these two additional notions is not offered, Millikan's account of REF as the foundations for proper functions remains incomplete. However, before we present such an analysis, two remaining problems need to be addressed. The first one concerns a suspicion of circularity in Millikan's explanation, the second touches on the need to loosen the definition of higher-order REF presented above.

Let us consider first the problem of circularity. One will remember that Millikan needed to introduce her technical notion of "reproduction" as a foundation for the establishment of the notion of REF. The notion of REF itself is supposed to help providing a theoretical account of "proper functions", Millikan's own elaborate version of the kind of teleological functions in respect to which a device could be said to successfully fulfill its purpose, or fail to do so. Is it then legitimate to make use of the notion of "proper function" for defining REF and then to use REF as a foundation for proper functions?

Attention should be paid to when the notion of "proper function" has been first introduced and to how it is used after it has been introduced. For the notion of proper function does not appear in the definition of first-order REF but only in the definition of higher-order REF. At this later stage, we already have a basic definition of REF in our hands that does not depend on any appeal to the notion of "proper function". Some proper functions are then identified in relation to first-order REF first. After that, the definition applies recursively. More generally, because the different aspects of the causal-historical phenomenon she is trying to capture interact through time in complex ways, Millikan's definitions are necessarily entangled but her explanation is not circular.

The second point to be cleared concerns the need to loosen the definition of higher-order REF. Remember that the strategy of the teleologist is to refer to the purpose of a certain device, what the device has been selected for, as the standard by which the performance of such device is to be judged. For that reason, the teleological notion of function is not reducible to a standard functionalist notion based on power or disposition. As a result, the definition of higher-order REF must be modified as to make sure not to exclude deficient devices which are incapable of achieving their purpose, a purpose they have been nonetheless selected for fulfilling. Hence Millikan amends her original definition of higher-order REF as follows:

> If anything *x* (*a*) has been produced by a device a direct proper function of which is to produce a member or members of a higher-order reproductively established family *R*, and (*b*) is in some respects like Normal members of *R* because (*c*) it has been produced in accordance with an explanation that approximates in some (undefined) degree to a Normal explanation for production of members of *R*, then *x* is a member of *R*. (Millikan, 1984, p. 25)

We have been describing the above definition as an amendment to the original definition of higher-order REF so as to make clear that this should not be taken as a new alternative notion. But presenting this amendment as a way of loosening the original definition is not entirely satisfying either. Such a description gives the impression that malformed items are granted the status of members of higher-order REF in a manner of courtesy, or metaphorically only. This is not the case.

Malformed items do not owe their place in their respective REF because they share with the other items of such a REF some family likeness but rather because such items causally-historically fall under the same function category, as different as they may appear when compared with the other members of the family. As Millikan rightly insists, borderline cases must not be confused with deficient ones. One can argue whether or not chess masters fall under the category of professional athletes, but chess masters are not deficient athletes. Conversely, it is by virtue of having the right expected function that an item is properly classified as a member of some higher-order REF, not by functioning as expected with respect to such a function. This is the philosophical intuition behind the connection that teleologists establish between (causal) history and norm.

We now turn to the analysis of Millikan's notion of proper functions. Such an analysis demands that the notion of "ancestor" be first clarified. We will spare the reader the technical details of Millikan's definition. Suffice to say that, if A and B are two members of the same first–order REF and B has been produced by (direct or successive) reproduction(s) of A, then A is an ancestor of B. Now consider A' and B', two members of a higher-order REF. If A' has been produced by A and B' by B then, it follows that A being the ancestor of B, A' is the ancestor of B'. Of course A' is also the ancestor of B' if both have been successively produced by A when it is the proper function of A to produce such items. These precisions being made we can finally consider Millikan's definition of proper function:

> Where *m* is a member of a reproductively established family *R* and *R* has the reproductively established or Normal character *C*, *m* has the function *F* as a direct proper function iff:
>
> (1) Certain ancestors of *m* performed *F*
>
> (2) In part because there existed a direct causal connection between having the character *C* and performance of the function *F* in the case of these ancestors of *m*, *C* correlated positively with *F* over certain set of items *S* which included these ancestors and other things not having *C*.
>
> (3) One among the legitimate explanations that can be given of the fact that *m* exists makes reference to the fact that *C* correlated positively with *F* over *S*, either directly causing reproduction of *m* or explaining why *R* was proliferated and hence why *m* exits. (Millikan, 1984, p. 28)

This calls for a number of explanations and comments. Notice first that, although REF cannot be assumed to always have proper functions, it is always in virtue of being a member of a REF that a given item can be said to have a proper function when it does have one.

The definition introduces C as a “reproductively established or Normal character”. This, I believe, should be understood as follows. One will remember that p_1, p_2, p_3, etc., the properties that an item B possesses as the result of being modeled from an item A by a process of reproduction (in the technical sense defined by Millikan) have been called “reproductively established properties”. C is called reproductively established “character”, by reference to this terminology. The expression “or Normal character” is added, I believe, because Millikan wishes to take into account the fact that the set of all the properties shared by all the members of a higher-order REF are not, strictly speaking, reproductively established properties, for technically, only the properties shared by all the members of first-order REF are.

The use of the adjective “Normal” with a capital ‘N’, which qualifies C, echoes the use of the expression “Normal explanation” that was part of the definition of higher-order REF. We can now provide a more rigorous account of ‘Normal’ in place of the intuitive meaning of such a notion on which our understanding of Millikan’s definitions has been relaying so far. Hence, a Normal explanation for the performance of a given function F, spells out how F has been typically performed each time F has been performed properly in the past. “Normal” is minimally normative and essentially historical and must not be confused with “normal” in the sense of ordinary, common, or within a statistical average. A Normal explanation for the performance of F focuses on the set of historical conditions under which F has been performed as to be beneficial in the past. Those conditions are Normal conditions for F’s performance. The use of the adjective *Normal* can be then extended to apply to other aspects of Normal explanations, for example to qualify C as a Normal character of the REF (R). Each time some x is qualified as Normal by Millikan, it is with the concern to prevent dispositional or statistical readings of x and with the will to impose instead the intended historical/normative understanding. In more recent writings, Millikan has abandoned the practice of capitalizing the “n” of the adjective “normal” in the context of the exposition of her view, but her technical interpretation of this adjective has remained unchanged, as should the reader’s understanding of it.

Now, (1) refers to “certain” ancestors of m, rather than simply to “the” ancestors of m, because it should not be expected that every ancestor of m has performed F. Remember that the definition of higher-order REF has been loosened so as to include malformed members which may not be capable of performing F. In addition perfectly well-formed ancestors of m may have not encountered the cooperative conditions necessary to perform F, or on the contrary, may have managed to avoid situations where the performance of F was called for.

The combining of (1), (2), and (3) provides an account of proper functions that seems to make biological organs paradigm examples of items endowed with such functions. Because proper bio-functions are paradigmatic cases of proper functions, Millikan decided to extend the term “biological categories” to cover indiscriminately all proper function categories. This is the justification for the title of her book *Language, Thoughts and Other Biological Categories*. While philosophically comprehensible, this move in retrospect, proved to be largely counterproductive tactically, certainly contributing in part to the kind of misinterpretation about the nature of Millikan’s project mentioned earlier.

Finally, the definition above is introduced as a definition of “direct proper functions”. This is to be understood by contrast to “derived proper functions” the necessary complementary aspect needed for a full account of proper functions. This aspect of Millikan’s theory will play a decisive role in her treatment of some of the problems faced by teleosemantics with respect to mental content. Derived proper functions, along with the related notions of “adapted functions” and “adapted devices” will be the subject of the next section.

3.4 Derived Functions and the Novelty Issue

Proper functions are often relational functions. This is typically the case with the behavioral proper functions of animals which are using their perceptual and cognitive systems to take advantage of the opportunities offered by the environment. One of Millikan’s examples of relational proper functions is the case of the mechanism that allows the chameleon to alter its skin’s color to match the color of the background and as a result to become invisible to predators. The relative simplicity of such an example makes it handy and I will use it too as a pedagogical tool.

However, it is important to notice that such an example is limited in two ways. First of all, it is by altering its own physical states that the chameleon adjusts itself to the surroundings and permits the skin-color-adapting device to fulfill its proper function. This is indeed one possible way to implement a proper relational function. This is by no means the only way. Relational functions are also often fulfilled by modifying the surroundings so as to fit the organism or thanks to a coordinated modification on the part of the organism and in the surroundings.

Secondly, as I will explain soon, the role of relational functions is best described as the one of establishing a relational structure of a certain abstract type. In the case of the chameleon, the proper relational function of the skin-color-adapting device is to create and maintain the relational structure same-skin-color-as-background for the chameleon. The abstract structure here can therefore be described as a “same-as” structure. One should keep in mind that “same–as“ structures are only one kind of abstract structures among many others that relational proper functions commonly help to create or maintain. Some of these alternative structures can be also much more complex and more articulate than what may be suggested by the chameleon example.

Millikan’s famous example of the bee dance illustrates nicely such a complexity. The particular pattern of the dance that a bee returning to the hive performs in front of the other bees is used by those bees to fly in the direction where the nectar is located. The internal device that leads the bee to perform its dance has the proper function to create a relational structure between the pattern of the dance and the nectar location, specifying the direction and the distance of the nectar from the hive, and, it seems, what quantity of nectar is available as well. As a result, when everything goes as expected, the internal interpretative device of the bees observing the dance fulfills its own relational proper function, leading the observers to fly in a direction and within a distance from the hive that is structurally adjusted to the angle displayed by the motions of the dancing bee.

The bee dance helps to better appreciate which features of the chameleon skin-color example are common to the production of any relational proper functions and which are not. It helps also to make a further important point. The proper function of a device or a trait is often to produce a series of cascading effects, each stage having as a result the proper function to produce the next stage, until final completion of the global task. Hence, in the bee dance example, a first relational structure is created between the dance and the nectar’s location, the proper function of which is to produce a second relational structure between the dance pattern and the direction and distance of the observing bees' flight. This second relational structure has the proper function of making the bees fly in the right direction and find the nectar.

Abstracting from the particular circumstances and contingent determinations of each concrete implementation, the relational structures created by the devices or traits responsible for fulfilling relational proper functions, appear stable and enduring through time. For example, although one can expect the honey bee dance to differ from one occasion to another, mirroring the changes in the location of the nectar from one time to another, the rules governing the abstract relational structure that two distinct dance performances entertain with their respective nectar locations remain the same. Each time a geographer is mapping a new territory he is likely to resort to the same kind of relational structure, using the same scaling standard and relaying on the same type of abstract isomorphic relations, while the maps themselves keep changing according to which part of the territory the geographer is mapping for the occasion. In the same manner one bee dance has to differ from another to feed the observing bees with updated information about the location of the nectar each time this location varies. But the condition for the information to be properly read is that it is always coded from one occasion to the other using the same dancing code.

Millikan's introduction of the notions of "adapted function", "adapted devices" and "derived functions" to which we will turn now, can be understood in part as a way to do justice to the complex interweaved relational structures, functions and effects that generally participate to the fulfillment of proper functions. However a deeper philosophical concern is at work here. Because proper functions are historically defined, the suspicion arises that Millikan's model may not have the resources to explain novelty.

One can easily foresee how such a concern will become even more pressing when dealing with the proper function of human cognitive mechanisms, considering the human ability to forge original ideas, to easily imagine situations far remote from any ordinary or past experience, to seek out surprising new facts and to react with inventiveness to brand new problems.

Hence, a tension seems to exist between the looking-backward perspective of the notion of “proper function” and the capacities of creation and anticipation demonstrated by the human brain. The tension needs to be alleviated. It is therefore necessary for the integrity of Millikan’s philosophical perspective that the true adaptive abilities of proper functions, with the many long reaching effects of their implementation, be fully acknowledged. This is not an easy task, for Millikan must show at once that her notion of proper function when carefully unpacked, does have the resources to account for novelty, but also that her teleological model does not lead to postulate a new proper function at work behind each novelty. The following considerations should help to understand the gist of Millikan’s strategy in responding to this double requirement.

Compared with the proper-functional devices of octopuses, rats or crows, to say nothing about human beings, the chameleon’s skin-color adapting-device, or even the bee dance producing device, display a poor range of effects and a limited degree of flexibility. Yet, they manage to exhibit some apparently creative powers. For example, a chameleon may manage to adapt quickly to a pattern of colors never encountered before. It is important to notice the enduring or recurrent nature of the relational structures involved in such simple cases of adaptation to new circumstances; in the present case the “same-skin-color-as-background” relational structure. One should resist the temptation of postulating the existence of new functions each time elaborate organisms are successfully coping with unexpected circumstances by displaying a set of impressive "new", "smart" or "creative" behaviors.

A closer look may show that the same old proper function used in ordinary circumstances is at play, just exhibiting another effect of its same relational nature, using the same relational structure in the same operative way. The new circumstances could be responsible for giving a chance to this rare, surprising and yet Normal effect to come under the light. This is a more conservative, but also a more reasonable assumption than the postulation that the organism itself has just exhibited a brand new functional behavior on the spot.[14] Hence, Millikan needs to show how, over and above the many variations attached to each particular implementation of a given proper function, the same type of relational structure is put to work. It is with such considerations in mind that the notions of "adapted" and "derived function" should be investigated. These notions do not bring with them a new kind of functions. They are just new tools to help describing in details how (relational) proper-function producers fulfill their goals by adjusting to the many variations in the conditions under which they operate.

As promised, the chameleon example will now play its pedagogical role by serving as a straightforward illustration for these additional notions. Hence the skin-

[14] An old scholastic tradition has sadly given to the layman's uncritical assumption that "no higher degree of perfection is to be found in the effect that the one already present in the cause" the respectable status of a metaphysical principle. But consider two different animations of John Conway's game of life. One grid of the game shows three spotlights at rest on their squares, blinking endlessly. The other grid shows spaceships and gliders firing at one another, planets being destroyed, and a multitude of entities interacting in all the possible ways, in an ever evolving tapestry. Yet in both cases a simple device functions as to maintain the same abstract dynamic relation between spotlights over generations, always implementing the same algorithm made up of the same three basic rules. The only difference between the two animations comes from the original position of the spotlights.

color-adapting device possessed by the chameleon fulfills its proper function, making the chameleon safe from predators, by altering the skin color of the animal so as to satisfy the "same-skin-color-as-background" relation. On each particular occasion, satisfying such a relation demands that the device matches the skin color with the specific color of the background, brown when the chameleon stands on a brown ground, green when the background is green, and so on. A general description of the structural relation between skin-color and background-color as a "same-color-as" relation captures the relational function of the skin-color adapting device. A precise description of the particular way this relation obtains in a given set up, for example by making the skin turn brown when the chameleon is standing against a brown background, captures "the adapted proper function" of the skin-color-adapting device.

The difference between these two functions does not merely result from the greater degree of generality of one description compared with the specific details contained in the other. Turning the skin brown is not, as such, the proper function of the skin-color-adapting device because it is not what this device has been selected for in Sober's sense of "selecting for" (explained earlier). Turning the skin brown is not an operative part of the historical set of sufficient conditions explaining the ability of the chameleon to escape from its predators. To put it in Millikanian terms, the skin being brown is not part of the set of Normal conditions figuring in the Normal explanation for why the chameleon possesses an internal device with such a skin-color-adapting proper function. But of course, given that the background is in fact brown, the brown color of the chameleon's skin results nonetheless from the adapted function of such a skin-coloradaptor. Being the result of such an adaptive function, the brown colored skin of the chameleon is logically named by Millikan in such a context an “adapted device”.

By getting a proper grasp on the phenomena captured by Millikan’s terminology here, one acquires a valuable insight about the way history and novelty can be seen by the teleologist as supporting each other rather than conflicting with one another. For example, any specific dance performed on a given occasion by a honey-bee returning to the hive has the proper function of guiding the other bees to the nectar. Notice that, considered strictly as an adapted device, this particular dance may be entirely new. Beely, an adventurous bee, may have stumbled upon a new source of nectar located in a direction and at a distance from the hive that no bee has never pictured in a dance similar to Beely’s dance before. Hence, considered strictly as an adapted device, this dance has no ancestors and therefore could not have been historically selected for. How such a dance could ever be described as implementing a proper function then?

This is where the notion of “derived function” needs to be introduced. We have to be extremely careful in our description of what is going on in the case of Beely’s dance in order to understand in what sense such a dance can perform a proper function for which it does not seem to have been causally-historically selected for performing. We start first with Beely’s internal dance-producing mechanism which has the proper relational function of producing an abstract relational structure between the pattern of the bee dance and the location of the nectar. This is always done according to the same set of coding rules. The fact that such a mechanism possesses such a proper function is unproblematic for it has been selected for and passed on over generations of bees in the required causal-historical ways. The discovery of a new nectar location triggers the dance producing mechanism, which, in respect to the given context, adjusts the particular pattern of Beely's dance so as to indicate the location of the discovery. As a result, an adapted device is produced, namely Beely's new dance. Being stimulated by the dance, the other bees fly in the direction of the nectar. Producing such a reaction on the part of the other bees is the "derived function" of this adaptive device that is Beely's new dance. Notice that the flight that the observing bees perform after watching Beely's dance is an adapted device of its own. Notice also that it is as new and original as the dance itself.

When properly analyzed, the derived function of Beely's dance is not a new function at all but just an over-reaching consequence of the relational proper function of Beely's dance-producing device. Describe this derived function by considering the adapted device that performs it and what such a performance effects.   Suddenly it looks as if a brand new function performed by a brand new device, Beely's original dance, has produced a brand new effect: the flight of the other bees with its unique pattern. Abstract from the actual context of the concrete conditions in which such a derived function is performed and consider the adapted function responsible for generating Beely's dance. Now the same abstract relational structure that supports all the adaptive functions responsible for all the dances ever performed in accordance with the proper function of the dance-producing device of Beely's species of bees emerges. This abstract structure, enduring through time, provides a unified explanation for all bee dance performances across generations, however original or unoriginal their respective pattern might be. Add a parallel description of the role of the device which has the proper function of guiding bees to the nectar by adjusting the direction and the distance of their flight to the pattern of other dancing bees: you have now completed the explanation of how adapted devices end up performing proper functions they have not directly been selected for performing. In other words, you have provided an analysis of the mechanisms behind the creative power of the derived functions of adaptive devices.

Hence, under the right circumstances, proper functions lead to adapted functions which themselves generate adapted devices endowed with derived functions. One can see how natural selection could end up generating all kinds of novelty, recycling the same old tricks, putting them to "new" uses. Notice that the tension between the historical-conservative notion of "proper function" and the novelties that actual fulfillments of proper functions end up generating is alleviated only when the description of functions and how they perform remains at a certain level of generality, a level at which the role of enduring or recurring relational structures can be perceived.

3.5 From Natural Signs to Intentional Representations.

In section 1.3, we introduced Millikan's notion of "locally recurrent signs". The pattern of dark spots of Domino's coat and the footprints of quails on the forest ground were offered as examples of such signs. It is important to realize that what makes natural signs of this sort recurrent signs is not the fact that the same natural sign is likely to occur on successive occasions. Rather, it is the fact that, each time a token of a given type of natural signs occurs, the same relational structure between this sign and what it is a sign of, is likely to be satisfied. Local natural signs often contribute to the identification of the things which produce them or the things with which they tend to co-vary or co-occur within a properly specified context. Recurrent natural signs may carry useful information about the presence of food or predators, and animals are sometimes designed or conditioned to take advantage of this information and to adjust their behaviors to the presence of such signs.

However, recurrent natural signs are not intentional signs or representations. This is so because mere natural signs are *not* produced *in order to* be interpreted or used by any living creatures. Different species of animals may have evolved to the point where they have managed to read such signs and properly react to them, but mere natural signs have not been selected for informing any leaving creatures of the presence of any particular state of affairs. This is why, strictly speaking, natural signs cannot be wrong or empty. When mere natural signs end up being misleading it is only as the result of some misinterpretation on the part of the signs readers. The fact that natural signs cannot be false directly follows from the fact that they are not representational signs. Because they do not represent, mere natural signs cannot misrepresent either.

By contrast with mere locally recurrent signs, intentional signs or what Millikan originally labeled “intentional icons”, are members of REF (Reproductive established families) endowed with a proper function. Intentional icons or signs are generated by producing devices the proper function of which (in Millikan’s technical sense of the notion) is precisely to produce such signs. Producing-devices generate those signs in such a way that they can be interpreted by a given type of consuming-devices. Hence intentional signs stand mid-way between a producing device designed for generating them and a consuming device designed for interpreting them. The mutual adjustment of these two devices (or types of device) to one another is the result of a common history of co-evolution and/or mutual adaptation.

The bee dance is an intentional icon produced in order to be consumed by other bees, which, as a result, are stimulated to fly in the direction of the nectar, the precise distance and location of which has been specified by the particular features of the dance itself. The co-presence of the dancing bee and of the group of the other bees observing its behavior, are Normal conditions for the dance to perform its proper functions. The producing device for the bee dance has the proper function to make the dance correspond to the world, here to the nectar's location, according to a certain mapping rule. More generally, intentional icons are coded so as to be read by the consumer-device and benefit the organism that possesses such a device. It is the job of the producing-device to produce signs which can be properly interpreted by the organism, that is, signs which are true as understood or interpreted by the consuming device.

In writings posterior to LTOBC, Millikan has emphasized the decisive role played by the consumer perspective in establishing the representational status of intentional icons. This explains why Millikan's view is commonly referred to as "consumer teleosemantics". However, this denomination, somewhat unfortunate, encourages a distorted picture of her account of representational content. For Millikan's contention is not that the representational content of intentional signs depends on some well-defined univocal invariant behavior on the part of the consumer interpreting those signs and reacting accordingly. If such were her position, teleosemantics would be confined to the elucidation of the inner representations of very simple organisms, capable only of the most rudimentary and limited types of behaviors. As far as birds or reptiles are concerned, to say nothing about human beings and their elaborate mental networks of beliefs and desires, nothing can be described as constituting the unique appropriate behavioral reaction to the reading of a given intentional sign. Millikan's contention is rather that the content of an intentional sign is to be elucidated in relation to the set of Normal conditions under which the consumer-device is supposed to perform its function. This is the set of conditions the presence of which has been historically responsible for helping former consuming-devices of the same type to fulfill their proper functions in the past so as to benefit the ancestors of the actual organism. In that respect, the explanation is consumer-oriented because the content of the representation concerns what the consumer needs the intentional sign to correspond to. But it is important to keep in mind that it is not for the consumer to determinate the nature of such content by putting it to a particular use. Much more needs to be said about what constitutes such Normal conditions and how they should be distinguished from the many conditions constituting the general background on which the presence of these Normal conditions depend. We will return to this question in chapter four.

Meanwhile, to further address the concern that the scope of teleosemantics may reduce to trivial cases of innate behaviors or other conditioned responses to basic sensory stimulations, one must also readily concede that mere representations of the sort offered in the bee dance example come short from constituting the most elaborate form of intentional representations. By contrast to the elaborate forms of intentional representations best exemplified by human thoughts, these more primitive sorts of representations are named by Millikan "pushmi-pullyu" representations because their role is indeterminately descriptive and directive. The bee dance, for example, both at once pictures the location of the nectar and triggers the bees to fly and find such a location. By contrast, more complex organisms typically make use of intentional representations, the proper function of which is uniquely to describe or to direct, but not both. This certainly accounts in part for the much greater flexibility of the behaviors displayed by such organisms.

The proper function of the producing-device of a descriptive intentional representation is to properly inform the representation's consumer that a given state of affairs obtains. It does so by producing a sign which maps onto this state of affair in such a way that the consumer could interpret this sign properly and perform its tasks adequately. Notice that the content of intentional descriptive representations is not determined by the nature of the behavior or task that the consumer produces as a result of reading the sign. Rather it is determined by what it is that the sign needs to correspond to for the consumer to be in position to perform its task Normally. It is important to remark that in order to successfully operate as representations, descriptive representations do not need to help performing some unique specific task, the way pushmi-pullyu representations help the bees to access the nectar. The function of a purely descriptive representation can be combined in inferences with other descriptive or directive representations in such a way that the job it performs will depend upon the presence of these other representations and be part of the performance of a sometimes more distant and more articulated task. This constitutes the key element in solving the flexibility problem.

Also, it is necessary that the tasks that the consumer manages to fulfill by properly interpreting descriptive intentional signs be beneficial (at least some times) to the producer as well. This is needed to explain how producing and consuming devices have been selected to cooperate. The same condition of mutual benefit applies in the case of directive intentional representations. In such cases however, it is the consumer's responsibility to produce the proper state of affairs in order for the intentional directive sign to be mapping properly. The consumer does so by obeying the producer's directive sign as a guide for bringing about certain changes in the external world. Beliefs are descriptive intentional representations while desires are directive intentional representations. Hence mental tokens have indicative or directive functions depending on whether they involve descriptive or directive intentional representations. By engaging in practical inferences human beings are capable of integrating these two functions into a coherent picture in which beliefs and desires are connected in systematic manners. As a result, goals are stated, means to reach such goals are reflected upon and eventually actions are taken accordingly.

Additional elements are introduced by Millikan in order to account for the gap between human intentional representations and the ones of less sophisticated organisms. The following remarks are not intended to provide a detailed description of these features with all their implications but mainly to introduce them and to connect them to what has been presented so far. The capacity to engage in mediate inferences presupposes the ability to identify some middle term as having the same representational value in different premises so that a logical conclusion could be derived. Typically, the same entity will present itself on separate occasions, in different times and locations and through different sensory modalities. Drawing inferences implies the ability to recognize something as being the same from one encounter to another despite the many possible variations in its way of affecting the senses. In other words, the capacity to engage in practical inferences, so as to integrate descriptive and directive representations into unified elaborate intentional representations, implies that human beings have developed strategies for tracking and re-identifying objects or substances.

The content of mental representations is not directly accessible or transparently present to the mind in some immediate and incorrigible fashion. Providing a positive account of the way the identification of mental contents occurs is therefore an extremely challenging task. Millikan's book, *On Clear and Confused Ideas* (2000), is entirely devoted to addressing this challenge by providing a theoretical account of substance concepts. In order to do so, Millikan introduces many innovative concepts and ingenuous arguments to account for our ability to grasp sameness of content and eventually for our capacity, at least in some occasions, to know what it is that we are thinking about.

Hence, while Millikan's theory of intentional icons helps to explain the fact that some signs manage to be representational, that is to be about something, her theory of substance concepts provides an explanation on how we track, identify or re-identify objects or substances. The theory also helps to make precise the nature of some of the basic ontological assumptions supporting the other major distinctive feature of human thoughts, namely their subject-predicate structure combined with the use of the negation. The fact that human thoughts are articulated in this way makes it possible, at least in some occasions, to detect when two incompatible predicates, or a predicate and its negation, end up being ascribed to the same entity in violation of the law of contradiction. This, in turn, explains the possibility to discriminate concepts that are governing the use of terms which actually map onto the world from the ones which do not. In other words, combined with some basic assumptions about ontology, this specific subject-predicate structure explains our ability to decide whether our thoughts (or concepts) are empty or not, and when they are not, what it is that they really are about. These questions will be developed in section 4.3.

I have now completed the analysis of the main tenets of Millikan's model of teleosemantics. When taking a step back from the many details of the model, Millikan's perspective can be seen as the careful articulation of two main ideas. The first idea is that representations are the result of intentional signs respectively produced and consumed according to the proper functions of producing and consuming devices tuned to one another other through evolutionary or learning processes of mutual adjustment. The second idea is that representational contents of descriptive representations are determined by the Normal conditions under which intentional signs serve their functions of guiding their consumers during the performance of their proper functions. Intentional icons can be said to generate misrepresentation only by reference to what it is that they are supposed to represent. This normative element of intentional representations cannot be overlooked without making the problem of error impossible to solve. The distinction between true representations and false ones cannot be supported by a mere appeal to ad hoc counterfactual situations carved out as to satisfy the semantic intuitions of causal/functional theorists. Rather, the phenomenon of misrepresentation must be understood as resulting from the failure of representational devices to fulfill their teleological functions.

CHAPTER 4: OBJECTIONS TO TELEOSEMANTICS AND REPLIES

Chapter three was organized around the idea that organisms are endowed with the capacity to represent the external world because they possess evolutionary selected mechanisms with teleological representational functions. In Millikan's model these teleological functions are proper functions of representational devices which are members of reproductive established families (REF). The mental representations they generate are the result of intentional signs respectively produced and consumed according to the proper functions of such devices, tuned to one another through an evolutionary process of mutual adjustment.

In that respect, teleological accounts of intentional representations, understood as Millikanian proper functions or otherwise, crucially separate themselves from standard functionalist accounts by relying on a notion of function that is defined historically rather than in terms of statistical average, power or disposition. Hence it is by reference to its purpose, which means by reference to what a representational device is *supposed to* accomplish when operating under Normal conditions (Millikanian sense of "Normal") that the actual performance of such a device is to be evaluated. Such a normative expectation finds its justification by tracing back the history of the ancestors of the actual representational devices, revealing what they have been selected for, which in turn explains the presence of a similar device inside contemporary organisms belonging to the same species. By understanding misrepresentations as resulting from the failure of representational mechanisms to fulfill their teleological function, teleosemantics answers the so-called 'problem of error' and more generally the challenge of misrepresentation in a non-circular, non-question-begging and yet fully naturalist fashion. Something, I have argued, that neither causal-informational nor computational-functionalist models could do adequately.

In the eyes of the opponent of teleology, though, the picture is entirely reversed. For according to them, it is precisely the historical character of teleological explanations that offers the strongest and more basic justification for rejecting teleosemantics as a misconceived and ultimately incoherent perspective. The idea that the role played by history in teleological explanations leads to inadmissible consequences concerning the ascription or non-ascription of beliefs and desires to well-functioning organisms (well-functioning being understood here in functionalist standard terms) has become common wisdom for many in philosophy of mind. This last chapter addresses such a concern which constitutes the main reason commonly advanced by opponents for resisting the adoption of teleosemantics.

Section 4.1, *The Logical Structure of the Objections against History,* provides an analysis of the nature of the argument against the historical dimension of teleological explanations and presents a set of considerations explaining why the replies offered by teleosemanticists have done little to limit the negative impact of such an argument. Section 4.2, *Conceptual Analysis versus Theoretical Definitions*, evaluates the teleosemanticists' claim that teleofunctions are theoretical definitions which cannot be dismissed on the basis of semantic intuitions prompted by the mere contemplation of imaginary scenarios described in thought experiments. The value of this line of defense of teleological explanations depends on the rejection of two misleading assumptions about meaning which according to Millikan have been inherited from the method of conceptual analysis and which continue to be influential in contemporary writings. These assumptions are respectively named by Millikan 'the seed assumption' and 'the one-to-one assumption'. Section 4.2 offers an analysis of Millikan's refutation of these assumptions. Section 4.3, *Millikanian Meaning*, presents a comprehensive account of Millikan's elaborate perspective on meaning; a perspective that remains free of the misleading assumptions just criticized. Such a comprehensive account supplies the theoretical foundation required in order to support the teleosemanticists' dismissal of the standard types of objections against the historical dimension of teleological explanations. Despite the fact that they do not constitute any conclusive arguments against teleosemantics, the Swampman story and other similar thought experiments remain influential in discouraging the adoption of such a philosophical perspective. Section 4.4 uncovers a particular type of fallacy contained in thought experiments of this kind.

4.1 The Logical Structure of the Objections against History

Claims about the supposed undermining nature of the historical dimension of teleological explanations are often presented as conclusive pieces of evidence against teleosemantics rather than as a provocative aspect of such an approach that would call for a critical evaluation. Such an attitude has taken the form of an argument from authority which appears even in articles conceived as non polemical pedagogical resources.

A telling illustration of this can be found in Ned Block's revised entry on functionalism in *The Encyclopedia of Philosophy Supplement* (1996). After carefully reviewing the different versions of standard functionalism and the specific problems and limitations attached to each of them, Block eventually envisages the possibility that a teleological elucidation of the functional roles of mental representations may provide a way out of such difficulties. But before it is even seriously pondered upon or simply properly described, the mere possibility of adopting a teleological approach to intentional content is dismissed off hands with the blunt comment that:

> A major problem for this point of view is the lack of an acceptable teleological account. Accounts based on evolution smack up against the swamp-grandparents problem. Suppose you find out that your grandparents were formed from particles from the swamp that came together by chance. So, as it happens, you don't have any evolutionary history to speak of. If evolutionary accounts of the teleology underpinnings of content are right, your states don't have any content. A theory with such a consequence should be rejected. (Block, 1996)

Block's discussion of teleology stops there. No explanation is offered for why exactly such consequences are fatal and justify that teleology be dismissed. This unsupported conclusion, as well as the reasoning by which Block's is lead to the rejection of the teleological perspective, are typical of the way teleosemantics is dealt with by the vast majority of its opponents. With basically no introduction to what is presented as a self-defeating position, the reader is encouraged to dispose of an entire philosophical perspective without even engaging in the task of seriously studying the different models and elaborate arguments developed in accordance with such a research program.

Block's swamp-grandparents story is directly inspired by Donald Davidson's famous article *Knowing One's Own Mind* (1987) in which Davidson imagines that a lightning bolt striking a dead tree in a swamp accidentally creates a perfect replica of Davidson himself. This is followed by a discussion about what would be the actual status and content of the thoughts and memories of such a creature, assuming that one were willing to concede that this creature would have any. As we will see, Block is not the only one to have been inspired by Davidson's Swampman. Many theorists have offered their own twisted versions of this imaginary scenario to argue against (and sometimes for) teleosemantics. Swamptigers, swampcows or yet other versions of instantaneous or more gradual replicas of actual organisms have been imagined along with all sort of different principles for their creation. As a result, Swamp-beings of one kind or another populate the literature on intentional content and mental representation the ways all sort of zombies populate the literature on qualia and consciousness. Each new story brings new subtle difficulties but the core of the argument presented by Block here and, under various scenarios, by many other authors as well, can be recaptured, I believe, in the form of the following deduction:

1. In teleosemantics, the representational capacities of an organism O are explained by reference to the teleofunction of a certain type of representational device inherited from O's ancestors. Such a type of device has been causally-historically selected for its capacity to generate such representations and by doing so, being beneficial to the survival of the organisms endowed with such a capacity.

2. It is possible to imagine a counterfactual situation in which an organism O*, functionally undistinguishable from O (that is, under a standard functionalist reading of the notion of function) would nonetheless be lacking O's Darwinian background, or for that matter, any phylogenetic and/or ontogenetic history.

C. From 1 and 2 it follows that, while incapable of distinguishing O from O* nor any of their complex behavioral dispositions, a consistent proponent of teleosemantics must nonetheless deny O* the possession of the beliefs and desires she is willing to grant to O.

The conclusion C necessary follows from premises 1 and 2, but C is an absurd position and therefore teleosemantics must be rejected as a misleading theory.

One must first be clear on what the argument is not about and therefore on what does not constitute the nature of the disagreement between opponents and proponents of teleosemantics. In particular it is important to notice that, if on the one hand, the brief rendering of the role of history in teleosemantics sketched in premise 1 can be unpacked by the opposite side to the satisfaction of teleosemanticists, giving them some assurance that their view is properly interpreted, and if, on the other hand, teleosemanticists themselves are willing, if only for the sake of the argument, to grant the possibility of 2, then there is no dispute over C.

Moreover, the situation described in C is not a blind spot in the teleological approach that opponents of the view should be praised for having uncovered. The fact that their view had such consequences was all along well understood by teleosemanticists who did nothing to hide the facts or muddle the issue. From the start, that is in LTOBC, Millikan (1984), for example, makes sure to present the problem “so starkly that readers may close the book!” (p. 93). She imagines that out of a cosmic accident a bunch of molecules coalesce to create the exact double of her reader and comments:

> Though possibly that being would be and even would *have* to be in a state of consciousness exactly like yours, that being would have no ideas, no beliefs, no intentions, no aspirations, no fears and no hopes…. This because the evolutionary *history* of the being would be wrong. For only in virtue of one’s evolutionary history do one’s intentional mental states have proper functions, hence does one mean or intend at all, let alone mean anything determinate. To the utterances of *that* being, Quine’s theory of the indeterminacy of translation would apply—and with a vengeance never envisioned by Quine…. Ideas, beliefs, and intentions are not such because of what they do or could do. They are such because of what they are, given the context of their history, *supposed* to do and of how they are supposed to do it. (Millikan, 1984, p. 93)

The disagreement therefore is not about C. Nor it is about how the argument, if successful, could lead to the rejection of teleosemantics. Both sides, I believe, could agree that what is aimed at by the critic is a reductio ad absurdum. Proponents of standard functionalism and other opponents of teleosemantics want to support the view that history has no decisive role to play in accounting for intentional representations. The negation of this position is represented by premise 1 in the argument above. The contradiction that appears in the conclusion is that a teleosemanticist who agrees that O and O* are strictly indistinguishable must nonetheless argue that O has some characteristics, namely thoughts, beliefs and desires that O* is lacking. The fact that a contradiction is so generated leads to the conclusion that premise 1 must be false. If premise 1 is false then opponents of teleosemantics are correct in claiming that history does not play a decisive role in accounting for intentional representation and that teleosemantics is a wrong-headed perspective.

I would like to quickly point out in advance the different lines of argument opened to someone willing to contest the validity of such a proof. This can be done by targeting the second premise and/or the conclusion of the deduction. While it is true that the teleosemanticist willing to grant premise 2 must concede that the conjunction of 1 and 2 leads to C, it is also true that she has strong reasons to resist premise 2 in the first place. One basis for opposing 2 would be to remind the critic that teleological models aim at elucidating the nature of mental representation understood as natural phenomenon actually occurring in the real world. Mere possibilities about alternative realities are simply irrelevant. This reply raises the difficult question of deciding at which point some imaginary scenario counts as a counterfactual situation virtually taking place in the same reality as the one explained by the theory and at which point such a scenario is best described as depicting an alternative possible world.

Another basis for resisting premise 2 is to deny that any legitimate conclusion can be drawn from the mere possibility of imagining something. The fact that it is possible to imagine something does not show that what is imagined is possible. Let us notice also that the fact that one is actively picturing oneself as imagining something gives no guaranty that one is successfully imagining such a thing or, for that matter, anything at all.

A second line of thoughts can be developed targeting the conclusion. Let us start with the obvious point that any argument undermining premise 2 will undermine the conclusion as well. But even if premise 2 and therefore C are granted, the conclusion that the teleosemantic analysis must be rejected can be resisted in the sense that one can still deny that C is clearly unsustainable. The position of the theorist described in C may be counter-intuitive, but that is not enough to conclude that it is incoherent. This opens the question of how seriously we shall take our intuition as a guide in such matters. Let us just notice at this early stage of our analysis that if it is the position of opponents of teleosemantics that intuitions must count as a reliable test of what is possible or not, coherent or not, then surely, proponents of teleosemantics would be justified in pointing out that the scenario depicted in premise 2 appears to them (and maybe to others as well) intuitively even less plausible than the theoretical stand described in C.

These different lines of counter-attacks have all been explored in one way or another by Papineau, Millikan, Dretske, Neander and other teleosemanticists. However, a survey of the literature on the topic reveals that, by and large, and with few noticeable exceptions, their replies have been scattered and unfocused, embedded in sub-sections of articles centered mainly on other issues, partial in addressing only selected aspects of the problem and rather unsystematic in their exposition. Meanwhile Swampman types of criticisms about the historical dimension of teleological explanation have continued to make a profound and lasting impression on the opposite side. Is it then that teleosemantics is really undermined by this kind of criticism and that proponents of such a view have been incapable of defending their position convincingly against it? Such a judgment would be clearly unjustified.

In fact, two reasons of philosophical importance, I believe, explain why counter-arguments are difficult to produce in a rigorous and systematic manner and why such arguments, even when perfectly sound, have little chance to make a real impact on the opponents and swing the pendulum in favor of teleosemantics.

Firstly, offering a defense of the decisive role played by history in relation to mental representations is made difficult by the elusive nature of the different elements involved in swampman-like objections such as thought experiments, appeal to intuitions or logically possibilities as well as references to folk psychological notions about what constitutes a thought, a desire or a memory. The exact nature of each of these elements is controversial and the actual role that they should be allowed to play in philosophical arguments is open to debate. Furthermore, these different elements, which possess neither clear definitions nor precise boundaries, tend to overlap and support each other so as to give Swampman type of arguments a rather inextricable texture. As a result, no obvious angle suggests itself for taking a firm grip on them and launching a complete and consistent refutation.

Secondly, and even more problematically, the status of thought experiments, mental picturing of mere possibilities or semantic intuitions, as well as the confidence placed in such notions in support of a philosophical argument, are going to be largely dependent on the way the nature of intentional representation is itself understood. One important result of the analysis developed in section 3.2, *The Price of Teleology*, helps to better understand the complexity of the problem. In that section, it has been shown that the teleological perspective was best understood as an additional feature, conceived as a complement to pre-established models of intentional content, rather than as a positive alternative account of mental representation. To that extent, critics of teleosemantics have some reasons to feel justified in addressing the question of the historical dimension of teleological functions without engaging in describing any of the positive account of mental content offered by different teleosemanticists.

However, using Dretske's indicator semantics as a case study, it has also been shown that teleology was not a perspective that could be added on the top of basic models of intentional content without having the effect of reshaping these models. For example, the addition of teleology to Dretske's indicator semantics, I argued, challenges Dretske's central ideas that true representations are essentially the result of nomic causal connections carrying information from its source to its receiver in a fully reliable fashion. This lesson about the impact of teleology applies more generally. In that respect, critics of teleosemantics wrongly assume that they are in position to adequately evaluate the effect of the historical dimension of representational functions by merely hypothetically considering the adoption of teleology while keeping their basic assumptions about the nature of intentional content unchanged.

In the light of such considerations one can better appreciate the intricacy of the debate between opponents and proponents of teleosemantics concerning the role of history in teleology. Hence, the Swampman type of argument makes use of thought experiments, appeal to intuitions and to the kind of linguistic expressions that one might be more compelled to use in describing unexpected situations, as ways to question the integrity of the teleological perspective. Meanwhile, the confidence that, in general, one should have in this method of arguing for or against any philosophical view depends greatly on what one takes to be the real nature of intentional representations to start with. This, in turn, crucially depends on whether or not one is willing to add a teleological perspective to one's model of intentional content.

Presented with the Swampman type of objection the teleosemanticist cannot engage with the task of arguing for teleology in a manner that will fulfill the theoretical expectations and semantic intuitions of her opponent without undermining her own account of the true nature of intentional content. She cannot simply dismiss the objection by pointing out that it is based on a specific type of argument which itself depends on an inherited pre-teleosemantic conception of mental representation that turns out to be misleading. Such a stand will appear to the critic as simply begging the question.

I take it that very early in her work Millikan had the suspicion that a painstaking debate on the historical dimension of representational functions was doomed to fail because she could already anticipate that a teleological treatment of the proper functions of representational devices will eventually develop into a conception of intentional content that will be orthogonal (rather than merely opposite) to the account of the dominant models. Millikan's willingness to alienate some of her readers in LTOBC and to move on with the exposition of her original perspective was at the time the only reasonable strategy. In producing the best developed model of a naturalist account of mental representation that embedded teleology from the start, Millikan could hope that the final result will be impressive enough to generate a philosophical conversion to teleosemantics.[15] Now that such a model exists and that the reader has been familiarized with its main features in the precedent chapter, it is time to return to the objection against the historical dimension of teleological representational functions with these new insights in mind. The remainder of this chapter will be devoted to this task, starting with a new look at thought experiments and the intuitions attached to them.

4.2 Conceptual Analysis versus Theoretical Definitions

A common reaction on the part of most teleosemanticists presented with the Swampman type of objection is to stop this way of arguing in its tracks by opposing up front the idea that an appeal to mere logical possibilities carries any weight against their views. Hence both Papineau (2001) and Millikan (1989) insist that they are theorists engaged in the task of elucidating the true nature of mental representations which, as such, do actually occur in the real world as the result of biological processes. In order to fulfill such a task, they are led to introduce diverse notions of teleofunctions in their respective models. Such notions are to be understood as theoretical definitions. This kind of definition must be carefully distinguished from the descriptive definition of the kind offered by theorists engaged in conceptual analysis, a totally different research program aimed at identifying the ordinary meaning of, or establishing the common criteria of application for, terms like ‘function’ or ‘thought’ or ‘mental representation’.

[15] This strategy proved to be successful. New proponents of the view have produced original and substantive contributions such as Karen Neander, Godfrey-Smith, Crawford L Elder, Mohan Matthen, Carolyn Price or Linda A.W. Brakel.

As will be explained soon, teleosemanticists display a variety of attitudes toward the general enterprise of conceptual analysis, from Karen Neander's cautiously qualified acceptance to Ruth Millikan's resolute opposition. We will put aside these differences for a brief moment to focus on the claim that teleosemantics is not offered as a contribution to the advance of conceptual analysis and that teleofunctions are theoretical definitions introduced for the purpose of elucidating the nature of mental representations, in a spirit akin with the physicist's identification of heat with molecular motion or light with electromagnetic radiation. Such an account of the actual status of teleosemantics will limit the impact of the Swampman thought experiment, with its recourse to logical possibilities, at least in two respects.

Firstly, one common way to argue against teleosemantics on the basis of the Swampman hypothesis, that is, by an appeal to some supposedly shared semantic intuitions, loses its strength entirely. Any argument to the fact that, when faced with Swampman, a majority of members of our linguistic community may be tempted to describe such a creature using the same vocabulary and resorting to the same linguistic expressions ordinary reserved for the members of our own species is perfectly irrelevant for deciding the accuracy or inaccuracy of the theorist's account of the teleofunctions responsible for the mental life of human beings on earth. The fact that one may be tempted to describe Swampman as having beliefs, desires, and memories has no more bearing on the question of the adequacy of the teleological approach to mental representation than the fact that one could be tempted to describe Swampman as having a brain or a liver will have any bearing on the question of the accuracy of contemporary neurophysiology or hepatology.

Secondly, let it be granted that the Swampman scenario offers a description of an alternative reality that truly constitutes a logical possibility, leaving aside the vexing question of what the notion of "logical possibility" could ever amount to in such a context. A possible world may exist in which entities are at the same time generated spontaneously and behave in ways which, even in the long run, prove by any layman's standards, largely indistinguishable from the ones of human beings. Such a world will represent no serious challenge for a theory which identifies the production of mental representations with the fulfillment of some teleofunctions. The existence of the creatures inhabiting such a world will no more question the soundness of teleosemantics than will, on Putnam's twin-earth, the existence of the XYZ substance (with dispositional properties that makes it difficult to tell it apart from the earth substance called "water") the soundness of the theory by which contemporary chemists identify water with H_2O molecules.

Hence, the rejection of conceptual analysis coupled with the correct understanding of teleofunctions as theoretical, rather than descriptive, definitions seems to protect teleosemantics from at least some objections: the ones triggered by semantic intuitions about what to say when confronting the supposed logical possibilities of Swampman and other similar thought experiments. Theoretical definitions are not tested through semantic intuitions prompted by the mere contemplation of imaginary scenarios.

The efficacy of this line of defense depends on the three following assumptions: that, in principle, teleosemantics benefits from being understood as an entirely distinct enterprise from the one of conceptual analysis; that in practice theoretical definitions can be established independently of any significant recourse to conceptual analysis ; that the fact that some of the results of the teleosemantic approach conflict with the ones brought about through conceptual analysis can be safely disregarded in favor of the theoretical definition of teleofunctions. Certainly Millikan (1989), who famously describes the general enterprise of conceptual analysis as “a confused program, a philosophical chimera, a squaring of the circle, the misconceived child of a mistaken view of the nature of language and thought” (p. 290), will have no problem endorsing these three assumptions. The reasons for Millikan’s strong rejection of conceptual analysis are deeply rooted in her philosophical perspective and will be analyzed and carefully discussed below.

Interestingly enough, Millikan’s position is not universally endorsed by teleo-theorists. Karen Neander, whose teleological model shares many features with Millikan’s, most notably the historical dimension called into question by the possibility of Swampman, champions the idea that, when properly conceived, conceptual analysis is a valuable intellectual exercise with an important role to play in teleological explanations. What are Neander’s reasons for adopting such a view and what consequences follow regarding the line of defense against the Swampman type of objection just presented?

It should be noticed, first, that Neander’s notion of “conceptual analysis” turns out to be much weaker than the traditional one, since she does not introduce it as a search for necessary and sufficient conditions for the application of concepts or linguistic expressions that would constitute their essential meaning. Neander (1991) understands the conceptual analysis aspect of her work, more modestly, as “a search for the criteria of application that people generally have in mind when they use the term under analysis, leaving the issue of meaning aside” (p. 171). While acknowledging that “the criteria of application that people actually use are often vague, shifting, highly context-sensitive, highly variable between individuals and often involve perceptual data of a kind that is inaccessible, at least through philosophical method" (p. 171), Neander offers nonetheless three reasons for persisting in the endeavor of conceptual analysis.

Firstly, we may expect the criteria of application to be much less illusive when the linguistic community under study is constituted of experts, as it is the case with scientists making use of technical terms embedded in well articulated theories. This point is particularly important for Neander whose work has been concerned with identifying the way the notion of 'function' is put to work by contemporary biologists.

Secondly, Neander's version of conceptual analysis is "an attempt to describe what people think they are referring to" while "a theoretical definition, in contrast, is (roughly) an attempt to describe the thing referred to" (p. 171). As a consequence, Neander argues that when one has reasons to believe that a term may turn out to be empty, failing to refer, only a conceptual analysis will offer a path to successfully engage into a debate by helping to establish what people have in mind when they make use of terms such as, let us say, 'witch' or 'phlogiston'.

Finally, Neander argues that theoretical definitions themselves partly depend on conceptual analysis, in a sense that a successful theoretical definition will capture the actual nature of the reference of some concept which itself has been identified, in first approximation, on the basis of the use of the term designating such a reference in a given linguistic community. Neander's idea here seems to be that the more educated the community will be the smaller the gap between its linguistic use of a concept as established by conceptual analysis and the actual nature of the reference as captured by theoretical definition will turn out to be. From such point of view theoretical definitions and descriptive ones must eventually stand or fall together.

The upshot of this disagreement with Millikan on the value of conceptual analysis is that according to Neander "we can gain no immunity from the standard objections to the etiological theory by insisting that we are only interested in theoretical definitions" (1991, p. 173).

As I intend to show below, much more than meets the eye is at stake in this disagreement between Millikan and Neander. Ultimately, a full justification for Millikan's rejection of conceptual analysis requires a proper understanding of her critical analysis of the implicit assumptions at the core of what she has dubbed the inherited perspective of "meaning rationalism". It is important that the discussion about Millikan's objection to conceptual analysis be presented under the proper historical light. A key insight from Millikan's reading of recent analytic philosophy, concerns the persisting and detrimental influence of a certain method for approaching philosophical problems inherited from what is commonly referred to as "the linguistic turn" in the analytic tradition. Within the intellectual framework resulting from such a turn, philosophy of language not only takes the central stage, but according to Millikan, is endowed with a status more or less equivalent to the one of first philosophy. Such a status manifests itself by a conjunction of different factors that tend to support each other. First, because the task of the philosopher is understood as essentially consisting in the clarification, classification and proper articulation of concepts, any question arising in any given branch of philosophy needs also to be turned into an analysis of the actual meaning of the terms used to shape the issue. It is assumed that to be properly handled, the problems, for example, of the nature of causation in philosophy of science or the one of justification in epistemology require a prior elucidation of what constitutes “the” meaning or “the” proper criteria of application for the term “causation” or “justification”.

Such a prior analysis is also an a priori one. The method for investigating what the members of a given linguistic group do actually have in mind when making use of such terms is through conceptual analysis. This type of inquiry is not directly concerned with the empirical data and the experimental work in which the complex and elusive nature of the actual process of causation (or yet the challenge of eventually producing a proper justification for a given scientific hypothesis), present themselves to the theorist. The recourse to conceptual analysis, as the preemptive method for investigating issues in other areas of philosophy, has contributed to make philosophy of language resemble first philosophy even more by providing a self justificatory strategy for its privileged status. For conceptual analysis was the method of choice for investigating philosophy of language itself, making this dominant branch largely immune from any substantive objections, the value of such objections being ultimately decided on the basis of a conceptual analysis to be conducted in philosophy of language.[16]

The belief, inherited from the linguistic turn, that conceptual analysis must be the method of choice for investigating philosophical problems depends on two misleading assumptions about the nature of referential meaning. Millikan names them the “seed assumption” and the “one-to-one” assumption. Following the seed assumption, meaning

[16] It should be noticed that in contemporary writings, a different account of the notion of “conceptual analysis” is sometimes given which differs from the traditional one. This is particularly the case with thinkers who, like Michael Devitt (2005), believe that there is no a priori. Devitt sees conceptual analysis simply as the articulation and generalization of folk theories of the world, an enterprise only indirectly concerned with the analysis of concepts.

is an act that is achieved, whether successfully or not, entirely within the mind itself and for which an active and adequate cooperation from the external world is not required. As Millikan explains, according to the seed assumption:

> The intentional nature of the act of referring has its source or is given its shape by the mind. The mind, or its contents, alone determines the *criteria* in accordance with which a reference succeeds or fails to be made. The seeds of reference, if not the flower, are always entirely within the mind. (Millikan, 2005, p. 123)

Complementing the seed assumption, although theoretically independent from it, is the one–to-one assumption, the idea that:

> A univocal term in a public language is associated with one psychological state common to all competent users. For referential terms, typically the idea has been that the same seed of reference or criterion for successful reference must be grasped by all of its competent users. (Millikan, 2005, p. 123)

These two assumptions give support to the method of conceptual analysis in the following way. The seed assumption justifies the method of introspecting one's semantic intuition for testing how appropriate or inappropriate the use of a given term appears when applied to counterfactual situations, with the hope to capture its true meaning. The one-to-one assumption justifies the application of conceptual analysis to public terms by providing a theoretical reason for assuming that there exist one well-defined description for any given term, the one which adequately captures the actual feature that is shared by the psychological states of all the members of the linguistic community who have mastered the use of the term under study.

In many ways, more recent developments in analytic philosophy constitute a rebellion against such an inherited tradition. In fact, one would think that both assumptions have been discarded by contemporary critics. The seed assumption is considered to have lost most of its influence now that meaning externalism has supposedly asserted itself after Putnam's twin earth and similar arguments have convinced philosophers that "meaning is not in the head." It appears to me, and maybe it is Millikan's conviction as well, that contrary to the seed assumption, the one-to-one assumption has not been the object of purposefully conducted attacks. This may be because this second implicit assumption supporting conceptual analysis has never been as clearly identified as the first. However, Quine's rejection of the analytic-synthetic distinction and his attack on the notion of linguistic meaning are presented by Millikan as potentially undermining the one-to-one assumption because they seem to imply that "no separable inference connections will be learned as one learns one's language" (Millikan, 2005, p. 125). One can see how this laconic remark could be extended to show how Quine's holism and his defense of the indeterminacy of translation, to the extent that they have gained a surprisingly large group of proponents, could have contributed to undermine the one-to-one assumption.

Millikan’s particular insight is to remark that while most of the illusions attached to linguistic idealism and the overestimated power of semantic intuitions to transparently reveal the relation of our thoughts to the world may have been discarded, the most insidious aspects of conceptual analysis as a method of philosophical investigations have survived and continue to contaminate contemporary writings.

Hence, for example, Putnam’s twin earth thought experiment, as analyzed by Millikan, does not deliver on the promise to offer a proper setting for meaning externalism by getting rid of the seed assumption. Putnam’s argument shows that even if Oscar on earth and Toscar, his twin brother on twin earth, when both talking about “water”, are in an identical psychological state, the contents of their respective thoughts remain distinct, since Oscars’ thoughts are about water, while Toscar’s are about twin water. It follows that the extension of thoughts of natural kinds, like water or gold, are not fixed in accordance with, nor generally require a proper intellectual grasping of, their defining properties. Putnam understands natural kind terms as indexical which, like Saul Kripke’s proper names, are rigid designators: their respective extensions are fixed by their actual relations to thinkers in this world, our world. Natural kind terms therefore do not denote, in each possible world, whatever referent satisfies a given set of properties or dispositions according to some a priori definition. One does not need to properly grasp the essence of each referent in order to successfully think about such a referent either. In that respect, Putnam certainly encourages us to endorse a realist attitude towards natural kinds. According to Putnam, what ultimately is needed for the thinker’s thought to have the proper content is to be in what may be called, following Millikan, the proper existential relation. To successfully think about a given kind, or individual, one needs to stand in the proper historical, spatial-temporal, causal relation to this kind, granted that such a relation can turn out, in some cases, to be quite mediated and indirect. How do such considerations help establishing that meaning is not in the head? For even if kind terms are thought of as indexicals rather than, let us say, definite descriptions, it remains that indexicals are semantic elements too, and therefore the question remains also of establishing, among all the possible candidates, which one constitutes the correct existential relation that connects any given indexical to its proper referent. In the end, Putnam takes it that an understanding of this relation is what is in the head, determining what is to count as the referent. Millikan remarks that, ultimately, Putnam is forced to resort to what constitutes the 'intended' referent of the thinker's thought as the criterion for distinguishing among the many candidates which one constitutes the proper existential relation.

These remarks echo the analysis conducted in section 1.4 on the failure of Jerry Fodor's Asymmetric Dependency Theory (ADT hereafter). I noted that Fodor himself constantly justifies the ADT hypothesis assumption that it is the nomic connection between 'horse' and horses that is the robust one, the others being only parasitic, under the pretext that, after all, 'horse' is supposed to mean HORSE. I then pointed out that: "A highly complex network of causal paths is turned into a univocally semantic relation by cleaning the informational channels from the noise resulting from the interferences of unwelcomed accidental/contingent causal connections. This strategy can only succeed if the meaning of HORSE has been already independently established, waiting to be used as an evaluative standard to which the result of the causal process of mental tokening is to be compared".

The problematic aspect of such a result is not only that, contrary to what is commonly advertised, as a result of Putnam's analysis, meaning remains in the head after all; it is also that the method for identifying such a meaning remains itself prisoner of the method of conceptual analysis since it relies on thought experiments and the kind of linguistic intuitions prompted by the contemplation of imaginary situations. While presented as positions that could potentially challenge the seed assumption, Putnam's or Fodor's perspectives are developed in accordance with a method of reasoning that provides them with a conclusive force only if one already implicitly takes the seed assumption for granted.

Once the persisting influence of conceptual analysis is noted, one has reasons to suspect that the impact of Putnam's arguments against the one-to-one assumption may be less drastic than one could have imagined at first. Putnam's notion of the division of linguistic labor certainly shows that different people may think of a given kind in quite different ways. Sometimes, as it is the case when we entirely rely on the expertise of better qualified individuals, it would seem that our way of thinking about a kind is through definite descriptions making reference to what these better qualified persons have in mind when they think about it, something we may not be in position of thinking ourselves. That being said, if it is with respect to each thinker's intended referent that the criteria for deciding the success or failure of the act of referential meaning is to be determined, then it has to be the case that the referent of my own thought is identified by me using my own criteria, as mediated or idiosyncratic as such criteria may be when compared with the ones of some other members of my linguistic community. On the other hand, if such criteria are really understood as the expression of my own individual intention, how can one account for the success of social communication and how can one explain overall agreement in judgments in ordinary situations?

As long as the seed assumption remains unchallenged and keeps influencing their perspective, theorists will be tempted to conclude that something like a shared public meaning has to be postulated toward which the intentions of the speakers (and hearers) of the linguistic community must converge before reaching out to the world, each time these members are successfully thinking of the same thing. Hence, Putnam introduces the idea of some stereotypes shared by a community of thinkers and, similarly, as explained in section 1.4, Fodor appeals to some common ability to mentally grasp 'the' higher-order property that makes a given entity a member of its kind. Thus, as an effect of the remaining presence of the seed assumption, the one-to-one assumption eventually reappears under the form of the postulation of some public meaning, which must be mastered by any member of the community who counts as a competent speaker of the language.

The problem can be generalized over and above the particular views of Putnam or Fodor in the sense that any meaning externalist theorist, including the most uncompromising direct-reference theorist, needs to account for the criteria establishing the nature of the relation between thoughts (and/or linguistic terms) and their intended referents. This has to be done without begging the question by simply assuming that the intended meaning of our thoughts is already given independently of such an existential relation. The mere contemplation of logical possibilities does not provide the theorist with some direct mental access to such an intended meaning. All the members of the linguistic community, as thinkers, are crucially dependent on the actual cooperation of the world we all inhabit and within which such existential relations do actually take place. This includes the theorist himself who cannot claim to make use of an entirely different process than the ordinary mental processes at work for the rest of us when conducting his studies.

4.3 Millikanian Meaning

Millikan's personal perspective on the relation between language and thought, in particular her account of referential meaning, must be understood as an effort to do entirely without the method of conceptual analysis by definitively turning her back to the seed assumption and the one-to-one assumption. Some clarifications need to be made before tackling the task of explaining Millikan's biological account of language and thought with respect to referential meaning. Aside from the intrinsic complexity of Millikan's view, additional complications in exploring it come from some important shifts which have occurred through the years in her terminology. For the sake of clarity and in order to ease the reader's understanding, I will make use only of the most recent version of Millikan's technical vocabulary, retroactively updating, as one may say, the terminology introduced in less recent works, notably in LTOBC. Some of these changes in vocabulary need to be explicitly brought to the reader's attention. Thus, the expression ‘’semantic mapping” will be systematically substituted for the original one, later rejected by Millikan, of ‘Fregean sense’. In the same manner, the term ‘conception’ will be systematically replacing the one of ‘intention’. Finally and despite the fact that Millikan herself has recently more or less abandoned this practice, I will persist in the habit of capitalizing the N of the adjective Normal as in “Normal conditions” or “Normal explanations" when required by the context, to remind the reader of the specific technical sense of this term in Millikan's teleological model.

One last word of caution on the topic of vocabulary: It is crucial not to confuse the evolution in Millikanian terminology, which results mainly from pedagogical considerations, with any substantive change in Millikan's philosophical perspective. Millikan's well-justified disowning of some elements in her own original lexicon concerns exclusively the poor choice of words and the confusions it may have produced in the mind of commentators. In no way such a move constitutes a reconsideration of the soundness of the philosophical notions that such technical terms were intended to name from the start. Although best captured by the new vocabulary these philosophical notions, a constitutive aspect of Millikan's perspective to which we turn now, have remained, I believe, basically the same through the years and should be approached accordingly. With these clarifications in mind we turn now to the explanation of Millikan's view.

In different parts of her work Millikan has developed very detailed and subtle analyses on the question of conventional practices which I believe offer the best path for introducing her complex perspective on meaning. Thus, these are what I take to be the main results of Millikan's reflection on conventions which are directly pertinent for our present discussion. Conventions consist in reproduced, that is, handed down, patterns, which are proliferating at least in part owing to the weight of precedent and not simply because they turn out to be objectively more efficient than other alternative patterns in performing some given function. Hence, it is not enough for a pattern to proliferate among the population, all the people, for example, adopting a similar well-defined behavior, for such a (behavioral) pattern to count as a convention. Independent paths of personal discovery or trial and errors may lead a large group of people to end up adopting a similar strategy when faced with a similar problem. In such situations, the behavior not being handed down from one person (or group of persons) to another and the common adopted pattern not being copied from previous similar patterns, this common practice will not constitute a case of conventional practice.

For example, I have noticed a tendency of computer experts fixing bugs on costumers' machines to gently press with their fingers on the opened box of the DVD players of the PC units they are working on in order to make them close faster than they would have by pressing the open-close button. Such a practice saves time while the damage endured by the player's mechanism remains limited enough to escape the costumer's attention when the computer is returned. Assuming that such a practice is discovered and adopted by each computer expert independently, then, while being common practice among them, it is not conventional practice. Furthermore, even if it turns out that this behavioral pattern proliferates among the employees of some computer shop, being learned from their boss, this will not be enough to declare it conventional. This is so because, as noticed above, to count as conventional, taught behaviors or transmitted skills require the weight of precedent as a partial cause for their proliferation. In our example the pattern of manually forcing DVD players to close has an intrinsic superiority (at least from the computer scientist point of view) over any alternative methods which, in itself, accounts for its proliferation quite independently of the way it is acquired and without any reference to some historically pre-established practices. Conventions, by contrast, even when they help performing a given function, often display an element of arbitrariness in the chosen mode of performance of such a function which can only be accounted for by reference to the weight of precedent. To take one of Millikan's examples, the use of chopsticks in the Orient sustains itself as conventional practice because people having learned to eat that way, chopsticks are massively produced and made easily available to the population, providing the concrete set up in houses and restaurants for such a practice to be handed down to the next generation, which in turn, helps maintaining the massive production of new chopsticks.

While the handed down element involved in the reproduction of conventional patterns helps to differentiate conventions from mere commonly shared behaviors, the presence of the weight of precedent as a contributing factor in maintaining a conventional pattern alive, helps preventing the identification of conventions with prescriptive rules. Conventions do not typically involve formal regulatory prescriptions based on some abstract principle to be recaptured and properly formalized by the theorist. As shown by the chopsticks example, a conventional practice usually sustains itself for very concrete and mundane reasons rather than as the result of rationally pondered considerations about what some abstract rule may prescribe, as would be the case, for example, when playing chess.

A final remark needs to be made to the effect that some conventions are coordination conventions: they help fulfilling the common purpose of a group of people by making possible the cooperation of the different participants working together to achieve a given task. Conventional coordination patterns of activity proliferate partially because they make cooperation possible and, at least in a critical mass number of times, successful. But here again, the proliferation of such conventional coordinative patterns usually does not involve the rational processing of some abstract prescription. It is more likely learned from experience and copied from other members of the community with the implicit assumption that these people are like us: while coping with the world, they too learn from their mistakes and tend to replicate what has worked in the past; an attitude found in a range of living organisms that exceeds by far the few species, or maybe the only species, to which theorists are generally willing to grant true rationality. (A species-specific property of human beings in that context may be that it is easier for us to learn from the successful behavior of others than it is for most animals.)

The crucial next step consists in connecting such an analysis of conventions to Millikan's perspective on language and showing how such a connection contributes to a treatment of referential meaning that remains free from both the seed and the one-to-one assumptions inherited from the method of conceptual analysis. Such a connection is established through Millikan's contention that a central function of the language faculty found in human beings is to make language conventions possible, these conventions being, for the most part, coordinative conventions between speakers and hearers with the function of making communication possible. Public language forms, from words to syntactic forms, are conventional patterns which proliferate and are kept alive because, more often than not, they achieve their purpose, which is to help making communication successful. The handed down process partially supporting the reproduction of public language forms and the weight of precedent partially responsible for their proliferation are in strict compliance with what, according to Millikan, are the distinctive marks of conventions.

In fact, public language forms constitute coordination-conventional patterns endowed with Millikanian proper functions as such proper functions have been theoretically defined in chapter three. Millikan calls these proper functions "stabilizing functions" because when successfully activated, the coordination-conventional language forms endowed with such functions need to satisfy the communicational interests of both the speaker and the hearer, in order to remain stable and continue to proliferate as the behavioral patterns supporting them keep being reproduced. In a given linguistic community in which such stabilizing functions have successfully established themselves, speakers and hearers cooperate in communication by adopting the public language forms endowed with such functions in roughly the same way and for the same reasons they follow all sorts of coordination conventions unrelated to language. Thus, the display and proliferation of these public language forms in communication are the result of adapted behaviors on the part of speakers and hearers, both having learned from experience that one often sees her interests served by asking what one wants or by complying with what one has been asked, thanks in part to the stabilizing functions of such forms.

Because they are coordination patterns, requiring the conventional contribution of both speaker and hearer, the stabilizing function of public language forms fulfill a purpose that cannot be simply identified with the purpose of either of the two participants involved in an act of language communication. Furthermore, when cooperating in such an act by respectively producing and responding to tokens of such public language forms, speakers and hearers routinely adopt conventional patterns of behaviors without typically being aware of contributing to solving a coordination problem and without engaging in a process of deciphering or reflecting upon one another's thoughts, purposes or intents. As a result, public language constitutes a very complex field of active coordination-conventional language forms which are fulfilling functions with a purpose of their own. For example, the stabilizing function of some language forms are descriptive, others directive, the purpose of the former being to generate true beliefs in the hearers, the purpose of the latter to obtain their compliance. The actual mechanism responsible for the survival and proliferation of public language forms through the stabilization of such descriptive, directive or yet other proper functions in the social context of a given linguistic community resembles closely, at least in some respects, the mechanism of selection of the communicative signs of other animals in the biological context of Darwinian evolution. Both tokens of public language forms and tokens of, for example, the bee dance analyzed in the precedent chapter, are members of reproductively established families (REF) with proper functions. It is the success of such proper functions in the past that causally-historically explains the survival and proliferation of these tokens' types rather than the particular physical forms that these tokens present, the physical form retaining an element of arbitrariness.

Also these proper functions, whether it is the stabilizing function of a given language form or the biological function of the bee dance, can be identified and properly theoretically defined independently of any reference to anyone's intention or anyone's reasoning on what actually are such function's purposes. To be technically precise, the stabilizing functions of a given token of a public language form constitutes the direct proper function of such a token according to its type, that is with reference to the REF of which it is a member. By contrast, whatever the producer of such a token intends to express constitutes a derived proper function of such a public language form. When a speaker makes a Normal use of a given public language form, that is when she is not speaking metaphorically, ironically or in order to deceive, the literal meaning supported by the stabilizing function of the token and the speaker's intended meaning coincide. The important point here is that stabilizing functions of public language forms are direct proper functions of speakers' utterances and hearers' responses. Ultimately what decides whether or not the tokens of two temporally or spatially distinct speech acts belong to the same type, with a common stabilizing function, is not the fact that they may be formally identical to one another, nor the fact that they have been uttered by speakers sharing a common thought, but rather the fact that they originate from the same historical lineage and as a result belong to the same REF.

The stabilizing functions of public language forms are a first constitutive aspect of meaning. Meaning so understood, cannot be recaptured by inquiring into the psychology of speakers and hearers making use of such coordination conventions. Here is found a first justification for the rejection of the seed assumption. This first aspect of meaning is not to be established either by looking for some prescriptive language rule to which, it would be (wrongly) assumed, all competent speakers of the language need to subscribe. Stabilizing functions are proper functions resulting from an historical-causal process of (social) selection, they are not intellectual abstractions grasped through the method of conceptual analysis. Yet to evaluate to what extent the seed assumption and along with it the one-to-one assumption have been truly discarded rather than simply displaced, waiting for us further down the road, it is necessary to pursue the analysis of Millikan's perspective. Beside the stabilizing functions of public language forms, two additional elements need to be introduced namely ‘semantic mappings’ and ‘conceptions’, which are the other constitutive aspect of meaning. Let us turn to semantic mapping first.

It should be clear that the analysis provided so far remains incomplete. Public language forms are, for the most part, coordination conventions which, it has been explained, proliferate thanks to the success of their stabilizing function in helping communication. But what makes such a success possible? The parallel with natural signs is here again enlightening. When operating under Normal conditions, honeybee dances can successfully communicate the location of the nectar to other bees (and at once triggers them to go and reach it), because such dances are intentional icons endowed with representational powers. This aspect of Millikan’s teleosemantics requires special care for it is often an object of misunderstanding. True, because they belong to the same REF, it is the common proper function of all the bee dance tokens to indicate the location of the nectar and to trigger the flying of bees to that location. However, being historically-causally endowed with such a proper function is not what makes bee dances representational icons. In a nutshell and without restating the detailed explanation given in the precedent chapter, it is the fact that bee dances map onto the world in accordance with a projection rule which guaranties the isomorphism between the elements of the dance and the state of affair to which they correspond that explains how bee dances can be successful and therefore survive and proliferate. Because they are representational icons with satisfaction conditions in the external world, bee dances are intentional icons for they can fail in meeting such satisfaction conditions and as a result be false or wrong, an intrinsic characteristic of anything that qualifies as intentional.

Similarly, public language forms acquire stabilizing functions by successfully helping the communicative cooperation between speakers and hearers, but the success of a given form as a stabilizing function is itself explained by the semantic mapping which, in each case, determines the satisfaction conditions for sentences of this form. The stabilizing function of a descriptive sentence like "the keys are on the kitchen table" performs its job of helping the hearer to find what he is looking for because of the semantic rule according to which the sentence maps onto the world and under the condition that the truth condition is met, that is assuming that the keys are really where the sentence represents them to be. And in the same way the proper function of the bee dance is not what makes it a representational icon, it is not the stabilizing function of a given public language form which accounts for the representational power of sentences belonging to such a form. Such sentences are representational for the same reason that bee dances are representational, namely because both are governed by semantic mapping functions. And in the same manner that the bee dance's representational power stands on its own, quite independently of any reference to the dancing bee's intention, the representational nature of sentences of a given language form can be identified and studied quite independently from the speaker's intentional attitudes which they happen on some occasion to express.

Over and above the similarities between these distinct kinds of intentional icons, signs like the bee dance and conventional public language forms also differ in some crucial ways. First, as noticed earlier, the proper function of the bee dance and other similar pushmi-pullyu representations is indeterminately both descriptive and directive, while the stabilizing functions of different public language forms are clearly distinct and specific, as demonstrated by the respective role of descriptive and directive sentences.

Also, the relation between the semantic mapping function of a given language form and its satisfaction conditions is quite complex compare to what is the case with natural signs. The semantic mapping function of a given sentence results from rules according to which transformations operated on such a sentence, while leaving its syntactic form untouched, makes its satisfaction conditions vary. This is how, for example, sentences like “John loves Mary”, “Mary loves John “or yet “Mary hates John”, which all share a common syntactic form, have different satisfaction conditions. But the reverse is also true, in the sense that, two sentences may have the same satisfaction conditions, while their respective semantic mappings differ, as, for example, in “smoking is not allowed in this room” and “do not smoke here”. Hence, while the semantic mapping of a given sentence determines its satisfaction conditions the reverse is not true, since different sentences with different mapping functions can often share the same truth conditions.

A third level of complexity needs to be introduced concerning the relation between mapping functions and stabilizing functions to the effect that stabilizing functions can vary while semantic mapping remains unchanged, as it is the case for example with the two following sentences “will you go back to France this summer?” and “you will go back to France this summer”. These remarks reveal the complexity of the many possible relations between stabilizing functions, mapping functions and satisfaction conditions. To limit ourselves to what directly concerns the present discussion, the complex relations between the different aspects of public language forms just described play a decisive role in explaining how Millikan's account of meaning remains free from the fallacious assumptions embedded in the method of conceptual analysis. It does so by providing an alternative answer to a set of well-known puzzles. It has been traditionally claimed that solving these puzzles requires that over and above public reference, some public sense be postulated in accounting for the meaning of referential terms.

A first puzzle concerns the problem of accounting for the informative value of identity statements. If only reference is shared by users of referential terms, how could any identity statement of the form A=B ever be informative? Millikan's models as described above has the resources to solve this puzzle without postulating the existence of some public sense that must be grasped by all competent speakers and be part of their common psychology. Adopting Peter Strawson's perspective while providing it with an original support based on her teleological account of stabilizing functions, Millikan (2005) explains that the proper function of an identity statement of the type A=B is to merge in the hearer's mind the content of two information folders to which so far this hearer had access only through independent paths, respectively by means of the words 'A' and 'B'.

In the same spirit, Millikan explains that the stabilizing function of linguistic public forms in which the operation of negation is applied to assertions about existence, like in 'A does not exist', is to convert in the mind of the hearer the information about A into an inactive file, turning 'A' into a pretend name. The stabilizing function of forms of positive assertion of existence, like in "A exists" is to make such a file active again. The important point here is that thanks to Millikan's teleological account of the proper function of public language forms understood as stabilizing functions, no ontologically suspicious entity standing for the public meaning of referential terms needs to be postulated in order to explain how and why expressions of the forms ‘A is B’ and ‘A exists’ are not merely tautological or vacuous or how expressions such as “A does not exist” are not nonsensical or simply contradictory.

In fact, such expressions have a special status, for while being intentional icons they are not representational ones, and contrary to what may appear, in the context of such expressions, the terms they contain are not representational either. This is so because, as we have just seen, the proper function of such language forms is not to help identifying the state of affairs in the world onto which they are supposed to map according to their mapping functions, it is rather to help merging, activating or inactivating concepts.

This is where ‘conceptions’, the third and last constitutive aspect of meaning according to Millikan’s model, comes into play. Conceptions play a decisive role in keeping Millikan’s perspective on meaning free from the one-to-one assumption, the misleading idea that there must exist one psychological state or process common to all competent users of any given univocal referential term. Conceptions constitute this aspect of meaning which unlike stabilizing functions and semantic mappings is not essentially public. The central idea here is that, to have a concept of something, whether it is of an individual, a kind, or a property, is to be able to (re)identify such a thing from one occasion to another through its various manifestations. This is typically done through different means by different persons or even by the same person in different contexts. The conception that a person has of a particular thing amounts to the sum of all the different ways by which such a person is capable of recognizing this thing and is able to avoid confusing it with something else. Some of these ways of recognition may be grounded in our biological make up and directly activated by our perceptual apparatus, or learned through previous experience, in which case the identification comes as a direct consequence of a certain perception or conjunction of perceptions. In other cases the components involved in a certain conception may lead to the identification through the explicit use of descriptions which involve the appeal to prior concepts.

In any event, no particular way of identification needs to be shared by all the competent identifiers of a given thing and no given way of identification has a privileged status that would make it 'the' proper criterion for identification either. None of these ways of identification provides the essential or ultimate meaning for the referential term or the linguistic expression commonly used to denote such a thing. Also, none of these ways of identification are infallible; they are just more or less efficient, depending on the more or less cooperative conditions in which the tentative identification occurs, on how many ways of recognition one possesses, and on how complementary or independent such ways may be in their modes of application. It follows that knowing the meaning of a word in the sense of being able to identify what this word stands for is always a question of degrees. It also follows that any additional knowledge acquired about a given thing becomes automatically part of one's conception of this thing, since this acquired knowledge potentially offers new ways for future identification.

The general question of how an individual cognitive system manages to implement mapping functions, recognizing things that are the same as the same when they are encountered again, so as to represent them consistently in thought is certainly a very challenging one, especially since, even in the case of direct sensory perceptions, the object to be identified remains distal and is accessed only through proximal stimulations of the senses which are subject to great variations from one encounter with the thing to another. The discovery of the actual mechanisms involved in the cognitive process of (re)identification is, for the most part, an empirical question to be advanced through scientific research. However, Millikan's perspective provides a promising philosophical framework for such a research.

Millikan's uncompromising externalism with respect to meaning implies a commitment to a realist ontology in which the world contains roughly bounded Aristotelian-like substances which are making up the fabric of the world quite independently of any linguistic rule or culturally established definition. An original epistemology of substance concepts completes Millikan's framework of metaphysical realism and meaning externalism. To limit ourselves to the aspects relevant to the present discussion, the three main kinds of substances distinguished by Millikan, namely historical kinds (like cats as a species), eternal kinds (like gold) and individuals, although for different reasons, all provide a fertile ground for inductive knowledge. Tokens of the same kind are not merely the various elements of a common set or class which have been classified as members of this set because they happen to share some interesting overlapping properties.

If kinds were mere classes, some given public criteria would be both necessary and sufficient for the identification of substances and the meaning of a substance concept would be truly mastered only by the individuals within the linguistic community who were capable of properly spelling out and testing for such criteria. Identifying or (re)identifying a substance would be an all or nothing affair rather than an ability subject to different degrees of mastery. Also, if kinds were mere classes, the content of substance concepts would be constituted by reference to some public meaning which itself would have to be both established and understood on the basis of other and supposedly more basic concepts, the meaning of which would have to be identified in the first place. But if the mastery of the (hypothetical) public meaning of a kind term is required for a successful (re)identification of the members of this kind, how the (re)identification of the different tokens of a given type of kind terms could ever occur in the first place?

Substance concepts are picking out real kinds, they are not merely classifying entities according to some public definitional criteria and they are not defined by the method, or rather the many different methods, by which we may identify them. The different members of an historical kind, the different samples of an eternal kind, or yet the different spatiotemporal slices of a given individual, being different tokens or parts of the same substance, share, in each case, many common features and do so for good reasons. For example, the causal-historical process responsible for the reproduction (in Millikan's technical sense of reproduction) of the different members of the same historical kind explains that one can acquire a great deal of valuable information about such a kind through an encounter with only few of its members. By dissecting one or two specimens, a student in biology can reasonably expect to acquire a general understanding of the nervous system of the frog as a species, not unlike the way one acquires, on first encounter, valuable information about a person, her size, the tone of her voice, whether she is right handed or left handed; information that will be safely, if not infallibly, used for future identification. Substances possess many properties that are clear and fairly reliable markers of their presence, at least in an adequate context. This explains how different persons can converge in judgments when identifying a substance and how such an agreement is reflected in their concordant use of the corresponding substance concept in public language, while sometimes very little overlapping, if any, is to be found in their respective conceptions of that substance.

The addition of a proviso about the adequacy of the context is required because the information gained from the presence of such properties, as is the case for any information gathered from the presence of natural signs, is always context-dependent for the reasons already examined in chapter two in reference with the quail tracks example. This leads us to the final original aspect of Millikan's view on meaning that needs to be addressed here. Grasping the meaning of a referential term is not typically a question of mastering some public linguistic description, some necessary and sufficient criteria which will eventually be clarified through conceptual analysis; rather it is the results of one's conceptions, that is the different ways which together constitute one's practical ability to (re)identify a substance in given suitable contexts.

Identification is a very mundane process highly dependent on the presence of the proper conditions, that is, on the cooperation of the external world. Developing the ability to identify a particular referent or type of referents requires the interpretation of certain relevant natural information about it. The carrying of information is a physical process which implies that the emitting source and the receiver are participants in the same causal net. It does not follow that one cannot possess a meaningful identifying description for a referent which one has never and maybe will never physically encounter, for one can develop an ability for identifying something without ever facing circumstances in which such an ability will have a chance to be exercised. Also, ways of recognition involved in conceptions may themselves include conceptual components, as it is the case when a referent is identified by a definite description. New abilities are often acquired as the result of the cooperation of pre-existing (sub)-abilities and conceptions are often developed in part by articulating conceptual components which one already has mastered on the basis of previously acquired (sub)-conceptions.

At this stage, Millikan's last original claim needs to be introduced to complete her account of conceptions. Millikan argues that one also commonly acquires the ability to identify a substance without physically causally interacting with it (or not directly) by recognizing occurrences of the referential term for such a substance in public language. Other person's speech utterances are natural information carriers which operate as a medium that is helping the hearer's identification of the substance itself. By recognizing referential terms occurring in speech, one perceives substances through the medium of language and, thanks to this medium, (re)identifies these substances in the same manner that one can recognize a celebrity by watching her on TV. Tracking referential terms in speech and, as a result, perceiving and identifying a substance through language constitutes an economical and faster way to develop a conception of such a substance when compared with the work generally involved in developing a conception of the same substance through direct physical interactions with it. The price to pay of course is that the identification of substances through language alone is generally far less reliable, but this amounts to a difference in degree rather than in nature, since conceptions, whatever constitutes their process of elaboration, are abilities and no ability is infallible anyway.

The idea that language constitutes an alternative medium through which one directly perceives the world may seem quite radical but it is a view supported by strong converging evidence coming from experimental studies in psycholinguistics. A growing number of contemporary cognitive scientists embrace the view popularized by Daniel Gilbert (1991) that the belief mechanism in humans is governed by a Spinozan rather than a Cartesian procedure. In other words, contrary to what has been traditionally assumed both by philosophers and psychologists (supposedly under the influence of Descartes' philosophy) comprehension and acceptance are not two independent aspects of the process of belief acquisition. One does not first comprehend a proposition then, in a second moment and by a different mechanism, adopts a particular attitude toward such a proposition by an independent act of volition in which one considers such a proposition to be true and endorses it or on the contrary one rejects it for being false. Rather, our belief mechanism follows a Spinozan procedure: when faced with a new proposition we comprehend it and accept it at once. Believing what you comprehend is, so to speak, the default position. Skepticism, doubts and eventually refutations, when they occur, are secondary and demanding psychological processes which, in fact, carefully monitored tests consistently show do not come easily to us. A variety of ingenuous experimentations have been conducted in recent years, which have consistently delivered supporting results in favor of the Spinozan model (Gilbert, Tafarodi, & Malone, 1993; Knowles & Condon, 1999). Comprehension and acceptance are a piece in relation to what is represented through direct sensory experiences as well as in relation to what is apprehended by reading printed material or by listening to oral claims. In that respect, Millikan is justified in presenting language as an alternative medium of perception.

Our analysis of Millikan's perspective on meaning is completed. We are now in position to return to our original disagreement between Millikan and Neander over the role and value of the descriptive definition offered by the method of conceptual analysis. What is to be said about Neander's remarks concerning the meaning of empty names? The method of conceptual analysis, she argued, was required for establishing the meaning of referential terms lacking any actual reference, words like 'witch' or 'phlogiston'. In such cases, she implied, meanings could not be established by an appeal to theoretical definitions specifying the nature of the referents of these terms, since such referents do not exist.

The proper way to handle the problem of the meaning of empty names in the context of the teleosemantic model proposed by Millikan is, I believe, as follows. First, it is necessary to clarify the similarities and differences between sentences and the referential terms they may contain in relation to their respective meaning. Sentences are members of public language forms with stabilizing functions, which map onto states of affairs in the world and by doing so, are meaningful. The semantic mapping of a sentence has satisfaction conditions and the state of affairs which satisfies such conditions, when such conditions are indeed satisfied, is named by Millikan the real value of this sentence. Referential terms also are members of REFs with stabilizing functions and they too have mapping functions, i.e. they map onto their respective referents when they help to compose satisfied sentences. It is intuitively tempting to wrongly assume that the meaning of a sentence is derived from the more basic meaning of its referents. However, considered on its own a referential term does not map onto anything and therefore does not possess any real value. It is only as a constitutive element of a sentence which successfully maps onto a state of affairs, for example as a constitutive element of a true descriptive sentence, that a referential term has a satisfaction condition that can be fulfilled by the real value of its Normal referent. Referential terms like names' tokens are members of REF with stabilizing functions, and such REF exist because a critical number of their members have been successful in helping identifying their referents in the context of complete sentences in the past. In that respect, if one wishes, one can insist that referential terms are endowed with a meaning of their own, that is, quite independently of any sentential context. Such a meaning consists simply in the fact that these terms are supposed to refer, in the sense of acquiring a real value, when used as contributive elements within the context of the semantic mapping of a given sentence. The primary correspondence relation, the one supported by the semantic mappings responsible for meaning, is between sentences and states of affairs in the world. The correspondence relation between referential terms and their referent is semantically secondary in so far as it remains dependent on the implementation of the primary one between sentences and states of affairs in order to operate Normally.

The situation is even more complex when referential terms occur in the context of sentences for which the stabilizing function is not to generate accurate representations by mapping onto a given state of affairs, but is to help the hearer or reader of these sentences to merge, activate or inactivate concepts in their own mental framework. When occurring in such a context, referential terms do not have real values. This is not quite the same as saying that they are meaningless, since their meaning still consists in their referring, that is, in the fact that as members of particular REF they are supposed to correspond to a given referent, whether or not the sentential context within which they occur makes it possible for such referential terms to Normally refer.

Finally, the proliferation of empty terms, like ‘witch’ and ‘phlogiston’, when they are not just mistakes, may sometimes be accounted for, on the basis of some social or psychological functions they may help to accomplish, rather than by reference to any mapping function. Empty terms have nonetheless a proper function but they are not representational. Their meaning is not accounted for in terms of semantic mappings but rather results from the combination of explicit conceptual components articulated in descriptions traditionally passed from one member of the linguistic community to another. Contrary to real referential terms, and despite superficial resemblances with regard to syntactic shape and the semantic role they seem to play within the context of negative sentences (compare: “there are no ghosts in Scottish castles” with “there are no air conditioners in Scottish castles”) empty terms supported by traditional descriptions do not correspond to real abilities to identify any referents. The most that can be said about them is that they possess a “public meaning”, in the sense that they are supported by some public conception. Not being anchored to any particular referent in the world through any mapping function, the meaning (in a weak secondary sense of meaning) of such terms is likely to be subjected to ongoing modifications and even sudden shifts. The existence of such an elusive and capricious meaning remains possible only because the explicit components involved in the public conception of empty terms are themselves truly referential and correspond to actual abilities to identify. Millikan’s teleosemantics implies a full-blood externalist perspective according to which having a meaning for referential terms is first and foremost a question of referring to something. Conceptions, public or private, have only a secondary role to play in the constitution of such a meaning. Elucidating what is meant by a term, whether this term is referential or turns out to be no more than an empty name, is not to be decided through an investigation about what competent users have in mind when they make use of such a term or by the discovery of some unique criterion for application of this term that would be psychologically shared by all such competent users within a linguistic community.

A comprehensive analysis of Millikan's perspective on meaning vindicates her critical attitude toward conceptual analysis as a method of philosophical investigation. The proliferation of language forms as patterns of coordination conventions endowed with stabilizing functions, along with the semantic mappings supporting the representational dimension of the linguistic expressions cast in such languages forms, when articulated with an adequate theory of conceptions, that is the methods by which a competent user of the language manages to identify the satisfaction conditions for such linguistic expressions within the context of a given linguistic community, provide a comprehensive treatment of meaning. Millikan's account of meaning remains free from both the seed and the one-to-one assumptions and provides the theoretical foundation for the teleosemanticists' dismissal of Swampman type of objections, to the extent that such objections depend on semantic intuitions about the supposed public meaning of terms like 'belief' and 'representation' understood as descriptive descriptions to be established by conceptual analysis.

4.4 Misleading Intuitions Dispelled

In the last section, I argued that swampman-like thought experiments rely on a misleading conception of the nature of referential terms as well as on the unjustified idea that conceptual analysis could be playing a central role in theoretical models of intentional contents of mental representation. The result of such considerations is that an appeal to semantic intuitions about what a certain linguistic community would say when confronted with far-fetched counterfactual situations does not provide any conclusive objection against teleosemantics. Furthermore, as explained in detail in previous chapters, there exist several converging reasons for rejecting standard causal-informational and/or functionalist models in favor of teleosemantics. Yet, by exposing what many regard as the counter-intuitive results of the teleological approach, the Swampman story proves highly influential in discouraging the adoption of such a philosophical perspective. This is the reason why I will concentrate my efforts in this last section on the task of dispelling such detrimental intuitions.

I will start by introducing a counter-thought experiment due to Karen Neander. While the impact of Swampman intuitions is to question the soundness of the historical dimension of teleofunctions, Neander's imaginary scenario is so devised as to generate the intuition that such an historical dimension is, on the contrary, indispensable:

> Suppose that there are no lions. Then suppose that half a dozen lions pop into existence, we know not how. Having stared at them in stupefied amazement for some time, we eventually begin to wonder about their wing-like protuberances on each flank. We ask ourselves whether these

> limbs have the proper function of flight. Do they? When we discover that the lions cannot actually fly because their "wings" are not strong enough, we are tempted to suppose that this settles the matter, until we remember that organismic structures are often incapable of performing their proper functions because they are deformed, diseased, atrophied from lack of use, or because the creature is displaced from its natural habitat (the lion could perhaps fly in a lower gravitational field). On the other hand, often enough there are complex structures that have no functions, for instance, the vestigial wings of emus and the human appendix. The puzzle here is where among these various categories are we to place lions' "wings". I contend that we could not reliably place them in any category until we knew or could infer the lion's history. And if we were to somehow discover that lions had no history, and were the result of an accidental and freak collision of atoms, they would definitely not belong in any of our familiar functional categories. They are not then dysfunctional either because disease, deformity, lack of use, or because they are exiled from their natural environment. All of these require a past. Nor did they once have a function that they have now lost. Without history the usual biological/functional norms do not apply. (Neander, 1991, pp. 179-80)

Several remarks are in order here. Firstly, it should be clear that it is not my intent, nor is it Neander's, to offer the lion's story as a conclusive argument in favor of the theoretical notion of function supporting the teleological perspective. The reasons examined earlier which have led me to be radically suspicious about arguments based on intuitions triggered by the mere contemplation of so-called logical possibilities apply here with the same force as they do with Swampman. The merit of Neander's thought experiment is not to establish the soundness of the theoretical notion of teleofunction; rather it is to cancel out the effect of thought experiments' intuitions to the contrary.

Secondly, there is a certain ambiguity in Neander's analysis of the problem presented by the lions' story which demands clarification. This ambiguity is reflected in expressions such as "where among these various categories are we to place lions' "wings"" and "they would definitely not belong in any of our familiar categories". From the perspective I am defending, on the basis of Millikan's model, the challenge presented by the protuberances on the lions' flanks is clearly not a problem of semantics. The difficulty does not come from the fact that one finds it difficult to decide whether or not to call such protuberances "wings." But the problem is not essentially epistemological either. It is in fact quite common to be faced with a device for which we do not know the causal-historical background and yet are able to guess the function right. This is often a subject of misunderstanding on the part of critics of teleofunctions. These critics systematically point out that the function of the heart to pump blood was correctly identified by Harvey long before anything was known about Darwinian evolution and the historical process of natural selection responsible for the presence of hearts in living organisms. This kind of remark is of no force against teleological analyses.

Teleologists are not committed to the epistemological view that no correct description of natural devices' function could ever be given without the pre-established knowledge of their Darwinian history. Cummins' functions, for example, are a-historical; yet as explained earlier (section 2.4) the function-analytical explanations offered on the basis of Cummins' functions often provide illuminating descriptions of the functioning of complex artifacts or intricate natural systems. However, Cummins' or other functionalist analyses assign a function to a device in order to capture some of its dispositions or behavioral patterns, which for a reason or another, interest the theorist. Standard functionalist functions are theoretical notions which are especially framed in order to satisfy the theorist's interest and are ascribed to certain types of systems or devices. Of course, more often than not, the theorist is principally interested in understanding how such systems or devices work and what they are for. Yet, a functionalist explanation must not be confused with a teleological account of a device's theory-independently owned proper function. The distinction between the two kinds of explanations remains even when, as it is expected to happen on numerous occasions, the standard functionalist explanation hits on some proper function, as it is the case with Harvey's account of the heart's pumping-blood mechanism.

Hence, it seems to me that from a Millikanian point of view, what decides whether or not Neander's lions have well-functioning or malfunctioning wings, or for that matter, whether or not the protuberances on the lions' flanks are wings at all, is entirely a question of whether or not such protuberances belong to a reproductively established family (more likely a Higher-Order-REF) of devices endowed with such a proper function. The question raised by the lion's story belongs only in a secondary manner to epistemology. It is first and foremost a question of ontology. In the case of the proper functions of biological devices, the mechanism by which REFs causally-historically establish themselves, and by which their members are reproduced (Millikan's sense of reproduction) and proliferate, results from the process of natural selection. The fact that such a Darwinian process underpins the REFs of biological items is central to the success of teleosemantics. It is this fact that ultimately makes possible a fully naturalist account of mental content based on the teleofunctions of representational devices shaped by this evolutionary process. That being said, artifacts have proper functions too. Hence, the central point here is not essentially about Darwinian evolution and what one knows or ignores concerning the role that such a process of evolution may have played in the production of Neander's lions, with their wing-like protuberances on the flanks. Such lions after all, may have been genetically engineered by aliens on the basis of principles totally unknown to humans. Whether or not Neander's lions have wings is not a question of how well the dispositions of such protuberances could be properly described by the same function that a functionalist will ascribe to the well-functioning wings of flying animals on earth. Rather, the question of the nature of the lion's wings-like appendices is settled by finding out whether or not such protuberances are endowed with a proper function that it is their job to fulfill. This fact, and this fact alone, settles the issue. Hence the question of deciding the actual nature of the lion's protuberances does not result from some epistemological limitations or from a clash of semantic intuitions. The problem here is that the way the lions' story is told there is simply no fact of the matter concerning the presence or absence of any proper function.

Of course one could always pressure Neander to flesh out her story so as to provide the causal-historical background necessary to settle the issue. In the strict context of the briefly sketched scenario Neander is offering, the best that can be said in the spirit of a teleosemantic analysis is that Neander's lions are memes which are members of a REF of fictional entities. These memes keep being reproduced from her original writings and proliferate, including on this very page, because they are fulfilling their proper function of generating counter-intuitions in response to the Swampman story. This last remark already hints at the argument I will now present to conclude this final section.

I propose to uncover the presence of a very specific kind of fallacy hidden at the heart of swampman-like stories which has remained so far unnoticed. Before exposing such a fallacy, some clarifications need to be made in order to avoid possible confusions. Hence, it is not my wish to argue in favor of radical skepticism with respect to the use of thought experiments in intellectual debates. Such skepticism seems to me largely justified but the justification for it comes from a certain understanding of the meaning of referential terms as developed earlier in this work. Such an understanding itself springs from teleosemantic assumptions which the reader should not have to share in order to be convinced by what follows. This is why, for the sake of the argument, I will follow common wisdom and grant thought experiments a certain number of virtues. Thus, thought experiments are said to be extremely valuable in helping us to get a better grasp on the overreaching implications of revolutionary scientific theories. Langevin's twins' paradox (1911), for example, has facilitated a better appreciation of the implications of Einstein's theory of relativity. Aside from their illustrative value, most of the prestige attached to thought experiments comes from the role they seem to play in actively contributing to scientific breakthroughs. Galileo's famous reasoning about the free fall of a system of two bodies, one heavy, one light, attached to one another by a string, illustrates such a phenomenon. Galileo's thought experiment helped refuting the

Aristotelian principle stating that the speed with which a body falls is directly proportional to its weight. This line of support for the positive role of thought experiments can be extended even further. In recent years, Tamar S. Gendler (1998; 2004) has argued that thought experiments possess a unique heuristic value which confer them an indispensable role in the development of science.

Finally, thought experiments must not be too quickly discredited for depending on the assumption that there must exist some a priori knowledge to which one could get mental access through the contemplation of mere possibilities. A more nuanced position can be adopted which does justice to the practice of thought experiments and does not carry such anti-naturalist implications. Hence, Michael Devitt (2006) offers to understand the intuitions triggered by thought experiments as judgments which are empirically theory-laden central-processor responses to the phenomena envisaged in such imaginary stories. These judgments are not based on some conscious reasoning and may sometimes be innate (which in any case does not necessarily make them a priori) but in most cases, result from a long and rich past experience. Devitt's approach seems to do justice to the evidential role that intuitions play in philosophy without having to postulate the existence of a priori knowledge; a postulation which Devitt (2005) strongly rejects. With all that being said in their support, it remains that nowhere in science, it seems to me, thought experiments are presented as conclusive evidence in favor or against any theory, the way they are sometimes offered instead of a really argument in philosophical debates.

Let us now turn to the fallacy which, I am claiming, undermines the Swampman scenario. This particular fallacy is difficult to clearly identify. Fortunately for us, and sadly for philosophy in general, more rudimentary versions of the same type of fallacy can be found in several other thought experiments which have been presented as decisive arguments and continue to have a considerable influence in their respective domains. This is for example the case with the zombie twin scenario on the basis of which David Chalmers argues in favor of dualism. Notice that in what follows, I will refrain from engaging in a debate over the philosophical issue at stake or the pertinence of the view defended by the author, to strictly focus on the fallacious character of the argument. Once the fallacy in Chalmers' thought experiment has been made clear, it will be easier to identify the more intricate version of the same mistake in the Swampman scenario. Chalmers presents his imaginary story as follows:

> So let us consider my zombie twin. This creature is molecule for molecule identical to me, and identical in all the low-level properties postulated by a completed physics, but he lacks conscious experience entirely…. He will certainly be identical to me *functionally:* he will be processing the same sort of information, reacting in a similar way to inputs, with his internal configurations being modified appropriately and with indistinguishable behavior resulting. He will be *psychologically* identical to me…. All of this follows logically from the fact that he is physically identical to me, by virtue of the functional analyses of psychological notions…. It is just that none of this functioning will be accompanied by any real conscious experience. There will be no phenomenal feel. There is nothing it is like to be a zombie. (Chalmers, 1996, pp. 94-95)

On the basis of his zombie thought experiment, Chalmers concludes that “consciousness fails to supervene on the physical” (p. 97) and goes on arguing in favor of a revival of dualism in philosophy of mind.

For the sake of clarity let me name Chalmers’ zombie twin “Zalmers”. Suppose I am engaging in a conversation with Zalmers in which I am asking him to imagine that he has a twin brother identical to him in every other respect but lacking conscious experiences. Zalmers may be confused by my request, not understanding exactly what it is that he is supposed to imagine in terms of the difference between himself and his imaginary twin. After all, it is assumed that Zalmers himself is deprived of consciousness. If Zalmers declares that he does not believe that he can imagine his zombie twin and that the best he can do is to imagine an exact replica of his person, then Zalmers is not functionally identical with Chalmers. For Chalmers certainly believes that he can imagine having a zombie twin and will claim so, if asked. The thought experiment proposed by Chalmers entirely depends on this assumption.

On the other hand, it may be that Zalmers too will answer that he can actually imagine having a zombie twin; he has already considered such a situation and often reflects on it. This should not come as a surprise, since Zalmers is psychologically identical with Chalmers. But then, one wonders what it is that Zalmers is imagining which makes his twin a zombie rather than a mere replica of himself, since the difference between Zalmers and his zombie twin is about having conscious experiences and since none of them has any. Notice that it is important to assume here that Zalmers makes an honest report when claiming that he can imagine having a zombie twin. This is a necessary condition if Zalmers is to be psychologically identical with Chalmers. Our philosopher may be confused about what he takes himself to be capable of imagining, but he is certainly convinced that he can imagine Zalmers.

Maybe at this point Chalmers will want to intervene and explain that, although Zalmers honestly believes that he can imagine his zombie twin, he is wrong in believing such a thing: he is in fact failing to do so without realizing it. In section 3.1, I pointed out, following Millikan's analysis, that in ordinary speech, purposive behaviors are usually referred to by employing "success verbs" rather than "trying verbs", leading such terms to be equivocally used so as to cover both succeeding and merely attempting mental acts. This is certainly the case for the mental process of imagining something. "Imagining" ambiguously means both actively trying to imagine something and actually succeeding in doing so. This is why Chalmers would be entitled to adopt the line of defense suggested above according to which Zalmers wrongly, although honestly, believes that he is succeeding in imagining his zombie twin. This is in fact the only position that will allow the thought experiment to remain consistent. But I will now simply ask: what can possibly prevents what is true of Zalmers from being true of Chalmers as well?

Let me now step back from the particular occurrence of the fallacy in Chalmers' thought experiment in order to abstract its general logical form. First, this is not a case of begging the question, for the argument does not merely assume what it is supposed to demonstrate. The fallacy lays in the fact that in order to be developed in a consistent manner, the method of arguing on the basis of the type of thought experiment used by Chalmers requires the satisfaction of some general principles. These principles end up being violated by the results vindicated by the Zombie scenario, if such a scenario is to have any argumentative force.

Hence Chalmers' justifies is confidence in the value of the Zombie scenario by explaining that:

> Almost everybody, it seems to me, is capable of conceiving of this possibility. Some may be led to deny the possibility in order to make some theory come out right, but the justification of such theories should ride on the question of possibility, rather than the other way around. (Chalmers, 1996, p. 95)

Further developments offered by Chalmers in resonance with the spirit of these remarks help to better identify the general principles which need to be granted for his thought experiment to successfully establish the irreducibility of consciousness. Such principles can be recapped as follows:

1. There exists a well-defined notion of 'logical possibility' to which one can appeal in arguing for one's view on the basis of zombie-like thought experiments. Any counter-factual situation which constitutes a genuine logical possibility counts as a positive argument in favor of the view, shifting the burden of the proof on the opponent's side.
2. Whether or not a certain scenario represents a genuine possibility is decided by considering whether or not it implies a contradiction. Anything that is not

logically contradictory is logically possible and can be legitimately used as the basis for an argument of this thought-experiment type.

3. The presence or absence of a contradiction is itself revealed by a test of conceivability. Contradictions make the situations in which they occur inconceivable. By contrast, what after careful examination is conceivable is logically possible and judgments of conceivability are Cartesian judgments transparent to the mind.

A close reading shows that Chalmers agrees that such general principles are required by his argument and believes that these principles hold true. The reader at this stage certainly realizes that the position I am defending in this work leads to a straightforward rejection of all and every one of these general assumptions. However the presence of the fallacy in Charmers' argument is not contingent upon embracing or rejecting such principles.

A consequence of these combined principles is to make "conceiving" a strictly success-verb, ruling out a trying-verb reading of the expression. Chalmers is confident that he is successfully conceiving his zombie twin since he cannot perceive any contradiction in his judgments. To conceive the possibility of zombies is to succeed in conceiving such a possibility, for as long as you conceive yourself conceiving a given possibility such a possibility is successfully conceived by you. Conceivability so understood secures the move from not detecting a contradiction to detecting the absence of a contradiction. A commitment to this line of thought is required by the twin zombie argument. Yet as demonstrated above the only way to develop the zombie twin scenario in a consistent manner is to assume that Zalmers honestly reports that he is conceiving a zombie twin while in fact failing to do so. If thought experiments do have the argumentative force that Chalmers claims they have, then the zombie twin experiment demonstrates that logical possibilities cannot be established on the basis of the conceivability test. The fallacy in Chalmers reasoning comes from the fact that the zombie twin argument succeeds only given a certain interpretation of Zalmers' intentional states. Under such an interpretation the act of conceiving logical possibilities (or anything else) is shown to be as fallible as any other mental act. In fact the zombie scenario requires that the conceivability test with respect to logical possibility fails in the case of Zalmers in order for Chalmers' view on the irreducibility of consciousness to be successfully established. This in turn directly contravenes the general principles conditioning a successful appeal to a thought experiment such as the zombie twin argument in philosophical debates in the first place.

I will now turn to the analysis of the more intricate version of the same fallacy occurring in the Swampman thought experiment. Hence, Swampman is physically and behaviorally identical to Donald Davidson but, contrary to the famous philosopher, possesses no phylogenetic or ontogenetic history. From a folk-psychological point of view, Swampman, being indistinguishable from Davidson, must be considered as having thoughts, something that functionalism easily accounts for but that teleosemantics must deny. So, the argument goes, the Swampman scenario reveals the inadequacy of the theoretical definition of function at work in teleosemantics. The reference to the causal-historical process of Darwinian selection puts an unnecessary, and in fact detrimental, constraint on the explanation of the intentional content of mental representation.

Neander's lions' story shows how this kind of intuition can be turned upside down. More importantly the reflection conducted on Neander's thought experiment was the occasion to carefully distinguish two elements in the teleological treatment of functions. Firstly, there is the question of the existence of a REF of devices endowed with the proper function of operating as wings to which the protuberances on the lions' flanks may or may not belong. The protuberances are wings only if they are members of such a REF, that is, only if they have been selected for flying as the result of a specific causal-historical process. Secondly, there is the claim that when devices endowed with proper functions are found in natural biological organisms, like wings on birds' flanks or representational devices in humans' heads, such devices belongs to REFs which have been selected for through Darwinian selection. The Swampman scenario confuses these two elements in its rejection of teleosemantics.

This point is better appreciated by comparing the Swampman situation with the case of a "replicant" in the classic science fiction movie *Blade Runner* (Scott, 1982). Replicants are artificially engineered adult humanoids with no past. In a conversation with replicants' hunter Rick Deckard, a young woman named Rachael discovers that she is herself a replicant. What Rachael has taken so far as constituting her past experiences is no more than a stock of short stories artificially implanted in her brain. The spider she remembers observing laying its eggs in the garden when she was a kid, her mother who passed away and of whom she keeps an old picture, are fake memories. The picture is a fake souvenir. Having a memory of your mother requires that a very specific causal-historical process links you to a particular person which makes you her child. Rachael has no memory of her childhood, the information to which she has access mentally or through the contemplation of the picture are not memories of her mother. They are not about her mother and they are not memories. As soon as she realizes her situation, Rachael breaks into tears. We identify easily with Rachael's distress. Here, folk psychology firmly stands on the side of teleosemantics. In the context of such a story, a standard functionalist must hold his ground and proclaim that since the artificially coded information in Rachael's brain operates within the economy of her mental life in a way functionally similar to an actual memory of her mother, such information actually counts as memory. The fact that this information has been produced by a process different from the one generally occurring in humans is irrelevant. Memories of her childhood shape Rachael's personality in the same way they do for any of us. Functionalists could argue that if as the result of some technological breakthrough replicants were to suddenly populate the earth, terms like 'memory', 'thought' or 'desire' will soon indiscriminately be applied to the memories of humans and to the pseudo memories of replicants.

The teleosemantic perspective I am supporting rules out any appeal to conceptual analysis in an attempt to show that the true meaning of the term "memory" necessary implies the kind of causal-historical relations to past events that is lacking in Rachael's case. The point to be stressed is rather that memories are the product of devices endowed with the proper functions of capturing, stoking and retrieving information. The presence of such devices fulfilling their job successfully often enough to benefit our ancestors in helping them to survive and proliferate, explains why we end up being memory-devices owners as well. It is hard to see how such memory-devices could have been selected for in the course of Darwinian evolution if they had produced, stoked, and helped retrieving information entirely disconnected from any real past events and therefore largely deprived from any value for guiding future action. Once this point has been clarified, a teleosemanticist of the sort I am supporting can agree that the artificial device implanted in Rachael's brain is a pseudo-memory device endowed with a proper function. Such a device has been causally-historically designed to fulfill its purpose of producing pseudo-memories in order to provide Rachael with the psychological background necessary to support a normal development of her personality. Ultimately the proper function of such a device is dependent on the intentionality of Dr Eldon Tyrell, the scientist who in a long series of tries and errors has conceived successive generations of always more advanced replicants, with Rachael being a specimen of the last and most advanced type. (In fact Tyrell has modeled the spider story and other scenarios implanted in Rachel brain after his own niece's memories.) Notice that such a teleosemantic interpretation makes it possible to maintain a clear distinction between misrepresentation and true representation of Rachael's artificial past. Rachael may fail to pseudo-remember the scene with the spider in the garden or may pseudo-remember some of the details of the scene inaccurately. This would be the case each time the pseudo-memory device itself fails to fulfill its proper function. Rachael's case therefore offers no objection against teleosemantics.

On the face of it, the Swampman scenario seems more challenging since the creature suddenly emerging from the swamp in place of Davidson has no past history of any kind, natural or artificial. However a closer analysis helps in correcting such an impression.

Thus imagine that I have gathered detailed information about Davidson's childhood from reliable independent sources. I am now engaging in a conversation with

Swampman, Davidson's instantaneous substitute, just after he has emerged from the swamp. I want Swampman to tell me the color of his first bicycle, the one he was so fond of riding when he was a kid. If Swampman cannot remember the color of the bike or even the bike itself, I will be encouraged to dig deeper into his memory or rather pseudo-memory. As the failures to properly answer questions about friends, books or travels will accumulate, I may start to regard the Swampman's occasional correct answers as mere accidents. If in addition to being incorrect, most of the answers were inconsistent with the questions or simply unintelligible, I may start to suspect that the physical resemblance between Swampman and Davidson is a pure coincidence misleading me in adopting an over charitable interpretation of Swampman's vocal noises in order to give them a meaning they do not possess. I may finally reject the idea that Swampman has any mental life.

At his point the functionalist theorist will rightly object that, in such a scenario and contrary to what is required by the thought experiment, Swampman is not functionally equivalent to Davidson. But imagine that in response to my original question about the color of the bike, Swampman correctly answers that the bike was blue. I have with me an old picture of Davidson riding his bike to confirm this fact. The picture was given to me by a friend of the philosopher who found it recently as he was going through his personal archives in preparation for our meeting. Davidson and his friend have been arguing for years about the color of the bike. Davidson was convinced that the bike was red but, as the picture finally reveals, his old-time friend was right in remembering it blue. I will then simply ask what status should be granted to Swampman's correct answer. If it is a memory, what is it a memory of exactly?

At this point the functionalist theorist may argue that Swampman would never give such an answer since Davidson will have never given it in the first place. This commits the functionalist to a very literal understanding of the property of being functionally equivalent which may not be compatible with the way functionalist explanations are usually developed, that is, by reference to broad behavioral patterns, abilities and dispositions. Also, under such a narrow interpretation, Swampman becomes infallible in is rendering of Davidson's psychology. As a result, the property of being functionally identical ends up conflicting with the notion of being psychological identical which it was supposed to support. For Davidson himself, as any of us, is psychological fallible, as the picture demonstrates.

The functionalist theorist may reasonably endorse a more standard interpretation of the expression "functionally identical" in agreement with the common use of the expression in functionalist explanations. According to such standard interpretation, Swampman is functionally equivalent with Davidson and, like Davidson, psychologically subjected to possible errors. Swampman's answer about the color of the bike constitutes such an error. Whichever way the functionalist wants to argue, the important point for us is that, in each case, what decides whether or not the Swampman's answer is correct is not what happened in Davidson's past. The soundness of the Swampman pseudo-memory depends entirely on its capacity to successfully simulate accurate Davidsonian memories of the past, not accurate memories of Davidsonian past. In that sense, Davidson is to Swampman exactly what Tyrell's niece is to Rachael. The only difference is that, here, the artificial device, the proper function of which is to make pseudo-memories match their original models, is constituted by the telling of the Swampman story itself. We, the readers, are covertly asked to join the story-teller in endorsing Tyrell's role. This is where the result of the analysis of Chalmers' zombie story comes to the fore. First, the notion of establishing that something is a logical possibility on the basis of a test of conceivability supporting Charmers' zombie supports the Swampman scenario as well. The reasons for rejecting such a misleading notion which have been developed above could therefore be reproduced here. The reader will remember also that in order to be developed in a consistent manner, the method of arguing on the basis of the type of thought experiment used by Chalmers requires the satisfaction of some general principles which end up being violated by the results vindicated by the zombie scenario. A more subtle version of the same type of fallacy is at work in the Swampman argument. In order for the Swampman-type of thought experiments to conclusively support standard functionalist against teleosemantics, it has to be assumed that:

1. From a folk psychology standpoint, there exists a reasonably clear level of elaborate behaviors and dispositions, which when displayed by any given entity (E) guaranties that such an entity be granted thoughts.
2. A analysis of (E)'s mental representations can be successfully given by identifying the content of (E)'s thoughts with their respective functional role, that is, with the dispositions of each thought to interact with other thoughts, sensory states and motor skills which all contribute to the overall economy of the organism's behavior.

3. It is possible to successfully imagine a situation in which an entity (E*) will not be the result of any causal–historical process and yet be functionally identical, and therefore psychologically identical, with (E).

These principles lead to the conclusion that despite the fact that they have not been selected for by Darwinian evolution, Swampman's pseudo-thoughts are rightly described as thoughts by folk psychology, since they are functionally identical with the ones of Davidson. However, the analysis of Swampman's pseudo-memory of the color of the bike shows that the only way to develop the Swampman scenario in a consistent manner implies the violations of such principles. The violation is necessary in order to guaranty that Swampman remains psychological identical with Davidson as the result of falling under the same standard functionalist description. A consistent development of the thought experiment with respect to Swampman's pseudo-memories requires an implicit reference to the teleological function of such memories. In the imaginary context of the Swampman story the fulfillment of the teleofunction of such pseudo-memories is virtually fulfilled by this artificial device that constitutes the narration of the Swampman tale by philosophers and other intentional story-tellers. This in turn directly contravenes the general principles conditioning a successful appeal to thought experiments such as the Swampman argument against teleosemantics.

CONCLUSION

Non-teleological models of the intentional content of mental representation divide into two main perspectives: causal-informational theories and functionalist ones.

Causal-informational theories understand mental content in terms of co-variations between mental states and the physical events in the external world responsible for causing such states. A central idea supporting this type of models is that the distinction between nomic correlations and contingent ones with respect to such co-variations must account for the distinction between true representations and misrepresentations. The phenomenon of robustness, that is, the fact that a given item preserves its specific meaning while occurring in a great number of different circumstances and through very different modes of manifestation directly challenges this central idea. Taking into account robustness also helps at once to uncover and question a more implicit assumption supporting causal-informational models, namely the idea that misrepresentation itself reduces to a phenomenon of misperception broadly construed.

The result of my analysis is that, at best, causal-informational theories could help explaining how misrepresentations are produced but not why they are misrepresentations. Such considerations explain why Fodor Asymmetric Dependence Theory fails to offer a satisfying answer to the challenge of misrepresentation faced by causal-informational models. Also, I presented the challenge of misrepresentation through an analysis of Dretske's indicator semantics, a model in which such a challenge is more easily comprehended. This gave me the opportunity to point out two additional limitations attached to casual-informational models. First the type to type relation between mental tokens and the entities triggering their tokenings, which supports the nomic relation responsible for the production of accurate mental representations, renders the identification of individuals problematic. Second, and more importantly, the co-occurrence of a particular sign with the thing it indicates or signifies typically persists only within a well-specified domain. In that respect, Dretske's or any alternative context-free theoretical notions of information are deficient.

Properly representing something is not the same as having similar well-produced perceptions of the same item on two different occasions. Rather, it is successfully tracking down and re-identifying the same item under very different circumstances, thanks to different means, from one occasion to another. Such a process occurs always in a well-specified context. There is nothing to representing in general. A representation is always the representation of a particular thing, by some given means and devices, in some well-established environment and for the benefit of a particular organism or entity. These considerations are not additional characterizations about dispensable elements surrounding the occurrence of representation. These are constitutive elements of the phenomenon of representation itself.

The idea that accurate representations do not reduce to properly-produced perceptions resulting from reliably channeled information, could be seen as making the case for the adoption of functionalism. In functionalism, computational role theories and conceptual role theories alike identify the content of a given thought with its functional role, that is, in terms of what such a thought accomplishes within the general economy of the mental life of the organism. A causal-functionalist model is to be preferred to a strictly functionalist approach since such an approach would lead to the unfortunate adoption of anti-representationalism. Two-factor conceptual role semantics, a moderately externalist approach appears to be the best candidate for taking advantage of the positive features of causal-informational and functionalist models while escaping their respective limitations. However two-factor semantics, I argued, could not deliver on its promises. It is not simply that the actual connection between the two factors could not be properly elucidated by the theory. As any other functionalist models, two-factor conceptual role semantics depends on a standard functionalist notion of function that is inadequate. In a nutshell, functionalism manages to provide a principled distinction between well-functioning and malfunctioning devices only by relying on an intentional reading of the notion of function which violates the constraints of a truly naturalist explanation of mental representation.

This general statement can be unpacked so as to reveal a set of distinctive issues. Firstly, the standard functionalist account of functions relies on the notions of dispositions and abilities. Without the help of an intentional reading on the part of the theorist, such notions remain ambiguous and prove to be unreliable for defining the extension of any given functional kind of entities. Secondly, the functionalist theoretical definition for a given kind of devices conflicts with the functionalist criteria by which a particular device is identified as belonging to such a kind. Philosophers and scientists have provided theoretical definitions for natural kinds of items as well as specific procedures by which membership to such kinds is to be established. A similar strategy fails when applied to the functionalist definition of functional kinds of devices and the specific procedures for memberships to such kinds. This results in part from the dispositional nature of the standard functionalist notion of function and in part from the phenomenon of multiple realizations.

The upshot of the analysis of causal-informational and/or functionalist models of mental representation is that such models overlook the normative dimension attached to the production of the intentional content of mental representation. As a result, explanations developed within these theoretical frameworks must compensate for the absence of such a normative dimension. They do so by ascribing to the representational process under study an external evaluation commanded by the theorist's expectations.

By contrast with the notion 'of functioning as' used in functionalist explanations, teleological functions are defined in causal-historical terms rather than in terms of statistical average, powers or disposition. It is by reference to its purpose, that is, by reference to what a representational device is *supposed to* accomplish when operating under Normal conditions that the actual performance of such a device is to be evaluated. Such a normative expectation finds its justification by tracing back the history of the ancestors of the actual representational devices, revealing what they have been selected for, which in turn explains the presence of a similar device inside contemporary organisms of the same species.

Contrary to the functions of functionalism, teleological functions are not merely ascribed to representational devices; they are the proper functions which such devices possess independently from any observer's considerations. Ruth Millikan's philosophy shows how a detailed model of teleosemantics can be rigorously developed on the basis of such teleological functions. Millikan's version of teleosemantics is generally poorly understood and, as such, subjected to unfounded criticisms. While it is true that part of the difficulty one faces in studying Millikan’s work results from the technical complexity of her model, I believe that, ultimately, the real challenge comes from the true originality and depth of her philosophical perspective. The innovative nature of Millikan’s thinking is best illustrated by her treatment of referential meaning which is based on an analysis of the stabilizing functions of public language forms understood as the proper function of coordination-conventional patterns. Such a treatment remains free from the seed assumption and the one-to-one assumption, the misleading assumptions inherited from the method of conceptual analysis. It is my hope that the analysis presented in this work will render Millikan’s perspective more accessible and will contribute to generate a better appreciation for its true philosophical importance.

BIBLIOGRAPHY

Anderson, L. (1990). The Driving Force: Species Concepts and Ecology. *Taxon 39*, 375-382.

Bauer, R. M., & Trobe, J. D. (1984). Visual memory and perceptual impairments in prosopagnosia. *Journal of Clinical Neuro-ophtamology, 4*, 39-46.

Block, N. (1978). Troubles with Functionalism. *Minesota Studies in the Philosophy of Science, IX*, pp. 261-325.

Block, N. (1987). Functional Role and Truth Conditions. *Proceedings of the Aristotelian Society, LXI*, pp. 157-181.

Block, N. (1996). *What is Functionalism? (a revised version of the entry on functionalism in The Encyclopedia of Philosophy Supplement, Macmillan).* Retrieved from NYU.edu: http://www.nyu.edu/gsas/dept/philo/faculty/block/papers/functionalism.html

Block, N. (1999). Inverted Earth. In N. Block, & O. G. Flanagan, *The Nature of Consciousness* (4th ed., pp. 677-693). Cambridge, Massachusetts: A Bradford Book/ The MIT Press.

Brentano, F. (1874). *Psychologie vom empirischen Standpunkt* (Vol. I). Leipzig: Duncker & Humblot.

Brown, J. (1998). Natural Kind Terms and Recognition Capacities. *Mind, New Series, 107* (426), 275-303.

Chalmers, D. J. (1996). *The Conscience Mind, In Search of a Fundamental Theory.* Oxford University Press.

Churchland, P. M. (1984). *Matter and consciouness: A contemporary introduction to the philosophy of mind.* Cambridge, MA: MIT Press.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain.* Cambridge, MA: MIT Press/ A Bradford Book.

Cram, M. (1992). Fodor's Causaul Theory of Representation. *Philosophical Quaterly, 42* (166), 56-70.

Cummins, R. (1975). Functional Analysis. *The Journal of Philosophy, 72* (20), 741-765.

Cummins, R. (1995). *Meaning and Mental Representation.* Cambridge, MA: MIT Press/ A Bradford Book.

Damasio, A. R. (1990). Synchronous activation in multiple cortical regions: A mechanism for recall. *Seminars in the Neurosciences, 2*, 287-296.

Davidson, D. (1987). Knowing One's Own Mind. *Proceeding and Addresses of the American Philosophical Association, 60*, 441-458.

Devitt, M. (2005). There Is No A Priori. In E. Sosa, & M. Steup, *Contemporary Debates in Epistemology* (pp. 105-15). Cambridge, MA: Blackwell Publishers.

Devitt, M. (2006). Intuitions. *Proceedings of VI. International Ontology Congress (San Sebastian, 2004). 5-6*, pp. 169-176. San Sebastian: V. Gomez Pin, J. I. Galparsoro, & G. Arrizabalaga.

Devitt, M. (2008). Resurrecting Biological Essentialism. *Philosophy of Science, 75*, 344-382.

Devitt, M., & Sterelny, K. (1999). *Language and Reality* (2nd Ed.). Cambridge, MA: MIT Press/ A Bradford Book.

Dretske, F. (1981). *Knowledge and the Flow of Information.* Cambridge, MA: MIT Press/ A Bradford Book.

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

Dretske, F. (1995). *Naturalizing the Mind.* Cambridge, MA: MIT Press/ A Bradford Book.

Ereshefsky, M. (1998). Species Pluralism and Anti-Realism. (U. O. Press, Ed.) *Philosophy of Science, 65* (1), 103-120.

Farah, M. J. (1999). Visual Perception and Visual Awareness after Brain Damage: A Tutorial Overview. In N. Block, O. Flanagan, & G. Güzeldere, *The Nature of Consciousness* (pp. 203-36). The MIT Press.

Farah, M. J., & Wallace, M. A. (1991). Pure alexia as a visual impairement: A reconsideration. *Cognitive Neuropsychology, 8*, 313-334.

Field, H. (1977). Logic, Meaning and Conceptual Role. *Journal of Philosophy, 69*, 379-408.

Fodor, J. A. (1984). Semantic, Winconsin Style. *Synthese 59*, 231-250.

Fodor, J. A. (1985, Spring). Fodor's Guide to Mental Representation. *Mind*, 55-97.

Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind.* Cambridge, MA: MIT Press/ A Bradford Book.

Fodor, J. A. (1990). *A Theory of Content and Other Essays.* Cambridge, MA: MIT Press/ A Bradford Book.

Fodor, J. A. (1994). A Theory of Content, II: The Theory. In J. A. Fodor, *A Theory of Content and Other Essays* (3rd ed., pp. 88-136). Cambridge, MA: MIT Press/ A Bradford Book.

Fodor, J. A. (1998). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (4th Printing ed.). Cambridge, MA: MIT Press/ A Bradford Book.

Fodor, J. A. (1999). Psychosemantics or: Where Do Truth Conditions come From? In W. G. Lycan, *Mind and Cognition.* Blackwell.

Fodor, J. A. (2008). In Bensadoun, S., & Faye, P. (Directors). (2008). *Les Vues de l'Esprit* [Motion Picture].

Fodor, J. A., & LePore, E. (1992). *Holism: A Shoppers' Guide.* Blackwell.

Gendler, T. S. (1998). Galileo and the Indispensability of Scientific Thought Experiment. *The British Journal for the Philosophy of Science, 49* (3), pp. 397-424.

Gendler, T. S. (2004, Decembre). Thought Experiment Rethought—and Reperceived. *Philosophy of Science, 71*, pp. 1152-1163.

Gilbert, D. (1991, February). How Mental Systems Believe. *American Psychologist*, pp. 107-119.

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You Can't Believe Everything You Read. *Journal of Personality and Social Psychology, 65* (2), pp. 221-233.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature.* Cambridge University Press.

Godfrey-Smith, P. (2006). Mental Representation, Naturalism and Teleosemantics. In D. Papinau, & G. Macdonald, *Teleosemantics* (pp. 42-43). Oxford University Press.

Greenberg, M., & Harman, G. (2005, September 1). *Conceptual Role Semantics*. Retrieved from Repository UCLA School of Law: http://repositories.cdlib.org/uclalaw/plltwps/5-16

Harman, G. (1987). (Non-solipsistic) Conceptual Role Semantics. In E. Lepore, *New Directions in Semantics.* London: Academic Press.

Hull, D. L. (1965). The Effect of Essentialism on Taxonomy: 2000 Years of Stasis. *British Journal for the Philosophy of Science* (15; 16), 314-326; 1-18.

Jackson, F. (1986). What Mary Didn't Know. *the Journal of Philosophy, LXXXIII* (5).

Jacob, P. (1997). *What minds can do: intentionality in a non-intentional world.* Cambridge University Press.

Knowles, E. S., & Condon, C. A. (1999). Why People Say "Yes": A Dual-Process Theory of Acquiescence. *Journal of Personality and Social Psychology, 77* (2), pp. 379-386.

Kripke, S. (1980). *Naming and Necessity.* Oxford: Blackwell.

Langevin, P. (1911). L'évolution de l'espace et du temps. *Scientia, 10*, pp. 31-54.

LaPorte, J. (1997). Essential Membership. *Philosophy of Science, 64* (1), 96-112.

Levine, J. (1983). Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly, 64*, pp. 354-361.

Levine, J. (1993). On Leaving Out What It's Like. In M. Davies, & G. Humphreys, *Consciousness* (pp. 137-149). Blackwell.

Mayr, E. (1976). *Evolution and the Diversity of Life: Selected Essays.* Cambridge, MA: Harvard University Press.

Mayr, E. (2001). *What Evolution Is.* Basic Books.

McAlister, L. L. (1976). *The Philosophy of Brentano.* Atlantic Highlands, NJ: Humanities Press Inc.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism.* Cambridge, MA: MIT Press/ A Bradford Book.

Millikan, R. G. (1989, June). In Defense of Proper Functions. *Philosophy of Science, 56* (2), pp. 288-302.

Millikan, R. G. (1992). Review of A Theory of Content and Other Essays by Jerry Fodor. *The Philosophical Review, 101* (No. 4), 898-901.

Millikan, R. G. (1993). *White Queen Psychology and Other Essays for Alice.* The MIT Press.

Millikan, R. G. (1995). *White Queen Psychology and Other Essays for Alice* (2nd ed.). Cambridge, MA: MIT Press.

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

Millikan, R. G. (2004). *Varieties of Meaning: The 2002 Jean Nicod Lectures.* Cambridge, MA: MIT Press.

Millikan, R. G. (2005). *Language: A Biological Model.* New York: Oxford University Press.

Nagel, T. (1974). What Is It Like to Be a Bat? *Philosophical Review* .

Neander, K. (1991, June). Functions as Selected Effects: the Conceptual Analyst's Defense. *Philosophy of Science, 58* (2), pp. 168-184.

Neander, K. (1995). Misrepresenting & Malfunctioning. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 79* (2), 109-141.

Papineau, D. (1984). Representation and Explaination. *Philosophy of Science, 51* , 550-572.

Papineau, D. (1987). *Reality and Representation.* Oxford: Blackwell.

Papineau, D. (2001, June). The Status of Teleosemantics, or How to Stop Worrying about Swampman. *The Australian Journal of Philosophy, 79* (2), pp. 279-289.

Pöppel, E., Held, R., & Frost, D. (1973). Residual visual fonctions after brain wounds involving the central visual pathways in man. *Nature 243*, 295-296.

Preston, B. (1998). Why is Wing Like a Spoon? A Pluralist Theory of Function. *The Journal of Philosophy , 95* (5), 215-254.

Ridley, M. (1989). The Cladistic Solution to the Species Problem. *Biology and Philosophy, 4*, 1-16.

Rosenthal, D. M. (1999). A theory of Consciousness. In N. Block, O. Flanagan, & G. Güzeldere, *The Nature of Consciousness* (pp. 729-53). The MIT Press.

Scott, R. (Director). (1982). *Blade runner* [Motion Picture].

Searle, J. R. (1992). *The Rediscovery of the Mind.* Cambridge, MA: MIT Press.

Searle, J. R. (1997). *The Mystery of Consciousness.* New York, NY: New York Review of Books.

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

Sober, E. (1984). *The Nature of Selection.* Cambridge, MA: MIT Press.

Sober, E. (1992). Evolution, Population Thinking and Essentialism. *Philosophy of Science, 47*, 350-383.

Sterelny, K., & Griffiths, P. E. (1999). *Sex and Death, An Introduction to Philosophy of Biology.* Chicago: The University of Chicago Press.

Stich, S. (1983). *From Folk Psychology to Cognitive Science.* Cambridge, MA: MIT Press/ A Bradford Book.

Van Valen, L. (1976). Ecological Species, Multispecies, and Oaks. *Taxon, 25*, 233-239.

Weiskrantz, L. (1986). *Blindsight: A case study and implications.* Oxford: Oxford University Press.

Weiskrantz, L. (1990). Outlooks for blindsight: Explicit methodologies for implicit process. *Proceedings of the Royal Society of London, B 329*, 247-278.