

MIND AND REPRESENTATION

Thesis submitted to the
Institute for Graduate Studies in the Social Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Philosophy

by
Nazım Gökel

Boğaziçi University

2013

Thesis Abstract

Nazım Gökel, "Mind and Representation"

There are various kinds of views in the history of philosophy that have been offered as an account for the nature of mind. We can classify those views under two main categories as monist and dualist metaphysics of mind. Although monist and dualist philosophers seem to be in a complete disagreement regarding the analysis of the mental and the material, most of them at least seem to agree on one simple feature of mind: Mind is, at bottom, a representing capacity. The difficult task, however, begins exactly here, because now we have to find an answer for the following relevant questions: what is the ground of mental representation? In virtue of what does something in the mind come to represent something in the world?

The primary target of this thesis is to find an answer for the question about the ground of mental representations by surveying some of the theories of mental representation in the history of philosophy. As far as I know, we have three main views in the history of philosophy that have been proposed as the ground of mental representations: (i) the resemblance view, (ii) the causal view, and (iii) Aristotle's hybrid view that combines the intuitions of the causal and resemblance views.

Following Aristotle, I believe that the fact that mind is part of nature does not mean that representation consists in a physical/material process. In understanding the nature of mind as a representing capacity, we should also look for the formal criterion for representation as Aristotle does in *De Anima*. In this context, the philosophical investigation should not only limit itself to the physical/material explanation, it should go further and give a formal/conceptual analysis of what is being represented and misrepresented in mind. This task, however, is very demanding and it really pushes one to practice philosophy as excellently as possible. Aristotle's rigorous practice of philosophy in *De Anima* is an example of this kind of excellence. I think we should take the same Aristotelian attitude when we deal with the very nature of mind as a representing capacity. Representation is a very complex intellectual capacity, and we just cannot explain it through a materialist functionalist method, the method that is employed by Dretske and Millikan. Representation, as Aristotle would say, is essentially/conceptually related to both perceiving and thinking. That is the nature of representation, which very much resists any sort of material reduction.

Tez Özeti

Nazım Gökel, “Zihin ve Temsil”

Felsefe tarihinde zihnin doğasını açıklamak üzere öne sürülmüş çeşitli türlerde farklı görüşler vardır. Bu görüşleri, monist ve düalist zihin metafiziği olarak iki ana kategori altında sınıflandırabiliriz. Zihinsel olanı ve maddesel olanı tahlil etme noktasında monist ve düalist filozoflar arasında her ne kadar tam bir fikir ayrılığı olsa da, birçoğu en azından zihnin basit bir niteliği konusunda aynı fikirde gibi görünüyor: Zihin, en temelinde, bir temsil etme yetisidir. Fakat asıl zorlu görev tam da burada başlıyor, çünkü bu sefer bununla alakalı farklı sorulara cevap vermek durumundayız: Zihni temsilin zemini nedir? Zihindeki herhangi bir şeyin dünyadaki bir şeyi temsil etmesine sebep olan şey nedir?

Bu tezin öncelikli hedefi zihni temsillerin zeminine yönelik bu soruya felsefe tarihindeki birkaç zihni temsil kuramını inceleyerek bir cevap aramaktır. Bildiğim kadarıyla felsefe tarihinde zihni temsillerin zeminine dair önerilen üç ana görüş vardır: (i) benzerlikçi görüş, (ii) nedenselci görüş ve (iii) Aristoteles’in nedenselci ve benzerlikçi görüşlerdeki sezgileri kaynaştıran melez kuramı.

Aristoteles’in izinde giderek ben de zihnin doğanın bir parçası olması olgusundan hareketle temsilin de fiziki bir süreçten ibaret olduğu sonucuna ulaşamayacağımıza inanıyorum. Zihnin doğasını bir temsil etme yetisi olarak anlarken, aynı zamanda Aristoteles’in *De Anima*’da yaptığı gibi temsil için formel bir kriter de aramamız gerekiyor. Bu bağlamda, felsefi araştırma kendini sadece fiziki bir açıklama ile sınırlandırmamalı, daha da ötesine geçmeli ve zihinde neyin temsil edildiğine ve neyin yanlış temsil edildiğine dair formel/kavramsal bir çözümleme sunmalıdır. Bu görev ise çok emek gerektirir ve bu konu üzerinde düşünenleri olabildiğince kusursuzca bir felsefe yapmaya zorlar. Aristoteles’in *De Anima*’daki titiz felsefe pratiği bu türden bir mükemmelliğe örnektir. Bir temsil etme yetisi olarak zihnin doğası üzerine düşünürken, kanımca, aynı Aristotelesçi tavırda olmalıyız. Temsil, çok karmaşık düşünsel bir yetidir ve bunu Dretske ve Millikan’ın yaptığı gibi materyalist işlevselci bir yöntem ile açıklayamayız. Temsil, Aristoteles’in söyleyeceği gibi özü itibari ile/kavramsal olarak algılamaya ve düşünmeye ilişkin bir şeydir. Temsilin bu doğası ise herhangi türden materyalist bir indirgemeye şiddetle karşı çıkar.

CURRICULUM VITAE

NAME OF AUTHOR: Nazım Gökel

PLACE OF BIRTH: Famagusta, Cyprus

DATE OF BIRTH: 17.01.1978

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

Boğaziçi University

Yeditepe University

DEGREES AWARDED:

Master of Arts in Philosophy, 2005, Department of Philosophy, Boğaziçi University.

Bachelor of Arts in Philosophy, 2002, Yeditepe University.

AREAS OF SPECIAL INTEREST:

Philosophy of Mind, Artificial Intelligence, Metaphysics, Ethics, Kant, Aristotle

PRESENTATIONS:

Gökel, Nazım. "Kant on the nature of the feeling of respect." Kant Mini-Symposium #5: Themes from Kant's Ethics and Political Philosophy. Boğaziçi University, Istanbul. 3 June 2006.

Gökel, Nazım. "A Shock in Kant's System of Pure Ethics: The analysis of Kant-Herbert Correspondence." Yeditepe University, Istanbul. 18 April 2007.

Gökel, Nazım. "Artificial Psychology, Functionalism and Mental Representation." World Conference on Psychology and Sociology. Queen Elizabeth Elite Suite Hotel & Spa, Antalya. 28 December 2012.

PUBLICATIONS:

Gökel, Nazım. "Multiple Representability and Nature of Representation"
Kocaeli Üniversitesi, Sosyal Bilimler Araştırma Dergisi, 2009/2, pp. 75-85.

Kocaeli Üniversitesi Matbaası.

Gökel, Nazım. "Kant'ın Saf Ahlak Sisteminde bir Sarsıntı: Kant-Herbert
mektuplaşmasının yakın analizi", Felsefe Logos, 32, Ocak: 2007.

Gökel, Nazım. "Artificial Psychology, Functionalism and Mental
Representation." (Forthcoming) Procedia-Social and Behavioral Sciences
(ISSN: 1877-0428)

TRANSLATION:

Gökel, Nazım. "Cinsiyet gerekli midir: tetkik (1976/1987)" Ursula K. Le Guin.
Felsefe Logos, 38, Şubat: 2010.

DISSERTATION:

Gökel, Nazım. "The Limits of Pure Ethics in Kant." MA Thesis, Boğaziçi
University, 2005.

ACKNOWLEDGEMENTS

First, I would like to begin with special thanks to each member of the philosophy department for their endless and invaluable supports during the PhD coursework and thesis writing period. Second, I owe special thanks to all my colleagues Özge, Tuğba, Omca, Emin, Ertürk, Zeynep, Nurbay, Pelin, Onur and Evrim for their help and support. I would like to thank Güven Güzeldere for opening a course and introducing philosophy of mind. I thank each member of my committee for their kind behavior to be a part of the committee, and their constant help and motivation throughout the writing process. I am greatly indebted to Murat Baç who, in his own “kind” and “direct” way, remarked on all of the mistakes I committed in the previous version of the thesis, and reserved time to listen to my thoughts and problems. I owe great thanks to Sun Demirli, who “literally” helped me to stand up again, in the midst of all troubles, and open my eyes, again and again, to show me the light at the end of this very dark journey. I owe millions of thanks to Chryssi Sidiropoulou for her constant support, motivation and suggestions. Without you, I could not be able to finish this thesis. I understood, thanks to her, that a nice thesis work must be the product of the efforts of “finding the best angle”, as Zorba would say, and being satisfied with a “better than nothing” approach, as Misail would say. I also thank to Bülent Gözkan, who, as my former and lifetime teacher, have all been monitoring my development and work in Boğaziçi University. I am thankful for all your training and motivation. I still keep your letter of recommendation in my “secret” drawer, and most probably I will keep it till the end of my life. Last, but not less, I am indebted to Stephen Voss in such a way that I cannot ever pay back either in this life, or, if exists, in some other life. I still remember my very first “mental representation” of you at the backyard of the faculty building in 2002, and your first encouragement even started from there. I thank, from the bottom of my heart, for your priceless guidance, throughout all these years, both in philosophical and life matters. I hope I can make you proud as being one of yours “good” students, if I can. I thank you for your training, patience and friendship. In particular, thanks for teaching me to walk in the darkness and not to be afraid of it and hope to hit something beautiful eventually.

I thank Demet Yuncu Evrenesoğlu, İlke Yiğit Mordeniz and Alper Türken for their comments on the previous versions of some of the chapters of this work. I thank Nurbay Irmak for his very illuminating discussion about the nature of software. I owe special thanks to Chris Eliasmith, Michael Strevens, Tim Crane, Fred Dretske, Frank Jackson, Jack Copeland, Stephen M. Kosslyn and Jerry Fodor for their kindness in writing back to me and answering my questions.

In addition, I would like to thank three best friends: Tanzer Yakar, Yavuz İnan and Alparslan Özunal.

And then comes the family. For me, there are two people who graduated from Zorba Life School, my two uncles: Mahmut and Ali Diğil. I thank you for your support economical-and-mental-wise. I thank my parents, Fikriye and Tahsin Gökel, for their incorruptible belief in me. I thank my siblings, İlyas, Ömer and Hacer Gökel for their endless support and motivation. There were times when I was feeling so low and depressed.

Thanks for my nieces and nephews, at those times, for giving me inexpressible energy and love. Zeynep, Nüveyba, Salih, Özgür, Betül, you are my life energy. And, lastly, as Braz Cubas would say, “Ao verme que primeiro irá roer a carne fria do cadáver do ladrão(s), que entrou na minha casa no ano passado e roubou tudo exceto o disco em que o arquivo de toda tese foi armazenado, dedico com saudosa lembrança estas notas tese”.

CONTENTS

CHAPTER 1: INTRODUCTION.....	1
Intentionality as a Representational Affair.....	2
The Problem of Mental Representation.....	6
Précis of Chapters.....	11
CHAPTER 2: ARISTOTLE'S THEORY OF SOUL AND MIND.....	20
Introduction to Aristotle's Soul.....	21
Sensation.....	25
Imagination and Thinking.....	30
CHAPTER 3: INTRODUCTION TO METAPHYSICS OF MIND.....	46
Behaviorism.....	46
The Mind-Brain Identity Theory.....	58
Functionalism.....	70
The Platonic Ideology.....	90
CHAPTER 4: FROM CLASSICAL FUNCTIONALISM TO POST- FUNCTIONALISM.....	99
Intentionality and Representation.....	99
The Idea of Representation.....	102
From Functionalism to Post-Functionalism.....	107
CHAPTER 5: DRETSKE'S ACCOUNT OF REPRESENTATION.....	117
Dretske's Theory.....	117
Two Chief Troubles.....	129
CHAPTER 6: MILLIKAN'S ACCOUNT OF REPRESENTATION.....	152
CHAPTER 7: NATURE OF MENTAL REPRESENTATION.....	174
The Task and Theories.....	175
The Trouble with the Frog.....	192
The Aristotelian Way.....	198
CHAPTER 8: CONCLUSION.....	210
WORKS CITED.....	221

I dedicate this thesis to my late grandmother Fatma Diğil, whose agent
intellect is still with my patient intellect...

“Beyond your mind, there is no other Buddha...”

CHAPTER I

INTRODUCTION

“Each sign by itself is dead; what gives it life?”¹

I am committed to the following simple philosophical thesis, which, as far as I could see, has been recurrently defended, (depending on how it is read) outrageously rejected and sympathetically revitalized again and again by many people in different periods of the history of philosophy: Having a mind is, at bottom, a representational capacity.² Dennett, in his own lucid style of philosophy, remarks, “Descartes doubted almost everything *but* this”.³ This thesis work aims to clarify the concept of

¹ Ludwig Wittgenstein, *Philosophical Investigations* (Oxford: Blackwell, 1953), §432.

² There are two different readings of this simple formulation in philosophy. I will explain them in a short while. As for the general idea here “mind is, at bottom, a representational capacity”, I should specify this claim in order to prevent any possible confusion. One may say that there are some mental achievements that do not seem to require a representational capacity (e.g., riding a bicycle). Learning how to ride a bicycle requires having a mind since it is, after all, a mental achievement. But, it does not require that the subject represents all the minute details about riding a bicycle. So, having a mind is a more complex thing than having a capacity to represent. Mind cannot be explicated only with a reference to representation. So, it seems that I have to qualify the general claim about the association of mind with a representational capacity. First of all, I am not trying to reduce mind to having a representational capacity. However, I am making the following point: if a being/entity has a mind, then he can have a representation. So, I am dealing, here, with mind from a very special perspective. Of course, mind can be understood in many different ways, but this aspect of mind, i.e., a representational capacity, is the most important element for mental beings, especially human beings. Consider the following line of thinking: As a being with mind, I can take any single object in my environment (e.g., a pencil, scissors, eraser) and assign a representational role to each one of them, and thereby I can produce a system of representation through which I can convey my thoughts and feelings, how I see the world, or I can talk about a scene from Hamlet. How could I ever do these things? The simple idea is the following one: I can make objects represent something for others and me, because my mind is capable of producing representations to itself. Due to having a representational capacity, I can convey my thoughts/feelings to other mental beings. I thank Murat Baç for his criticisms and suggestions.

³ Daniel C. Dennett, “Artificial Intelligence as Philosophy and as Psychology.” In *Brainstorms: Philosophical essays on mind and psychology* (Cambridge, Mass.: MIT Press, 1978), pp. 119-122. Dennett here says: “First, the only psychology that could possibly succeed in explaining the complexities of human activity must posit internal representations. This premise has been deemed obvious by just about everyone except the radical behaviorists (both in psychology and philosophy—both Watson and Skinner,

mental representation in the literature and find an answer for the question about the very nature of mental representations, namely, the question “what is the ground of mental representations?” I believe that without having an answer for this question, we would not get any progress in understanding the fundamental nature of mind, that is, its being a capacity of representation.

In this introduction, I shall provide a sketch of this thesis work. To this end, I will first begin with a short introduction to intentionality. Then, I will talk about two different positions in the history of philosophy concerning the very idea of mental representation. The first position is based on the claim that part of being a subject is to be able to represent the world. The second position, however, goes further and states that in order to represent the world, the mind must store mental representations and utilize them later in cognition and action. Then, in the second section, I will spell out the problem of (mental) representation. In the final section, I will talk about the précis of the chapters of this thesis work.

Intentionality as a Representational Affair

So much has been said about intentionality and representation in the history of philosophy. Especially between the 1980s and 1990s in the literature, philosophy of mind has taken a very powerful and incredible spin towards the route to explain intentionality and mental representation by means of a functionalist-naturalist framework. The

and Ryle and Malcolm). Descartes doubted almost everything *but* this.” The italics belong to Dennett.

central idea at those times was that a subject of mind could be intentional without being conscious.⁴ So, the big question between the 1980s and 1990s was the mystery of intentionality: How can our thoughts reach out and be about other things in the world?⁵ In particular, how can material states of mind, differently from other material states in nature, bear content/representation? Here, let's pause for a minute and give a short description of intentionality.

Consider any basic thought like the desire to go to school. In a very simple desire like this, one's mind may go through many relevant thoughts. I may think of the school bag in front of me, and I may believe that if I want to go to school, I should bring the bag to the school. I may consider whether I should put some notebooks for taking notes at the class next day. I may remember how enlightening and enjoyable was the previous class, especially remember the lecturer's interpretation of Gregor Samsa's change of mind/body as a reaction to the capitalist/totalitarian system. I may also calculate the travel expense from home to the school, how much it would cost me to go to school, whether I am prepared to accept the fact that on the condition of choosing to go to

⁴ Hugh Clapin, "Introduction." In *Philosophy of Mental Representation*, edited by Hugh Clapin (New York, Oxford: Oxford University Press, 2002) p. 1; Tim Crane, *The Mechanical Mind* (London, New York: Routledge, 2003), pp. 27-29.

⁵ Hugh Clapin, "Introduction," p. 1. Also see Hilary Putnam, *Reason, Truth and History* (Cambridge, New York: Cambridge University Press, 1981), pp. 1-2; Fred I. Dretske, *Knowledge and The Flow Of Information* (Stanford, CA: CSLI Publications, 1999; originally published: Cambridge, Mass.: MIT Press, c 1981); Jerry Fodor, *A Theory of Content and Other Essays* (Cambridge, Mass.: MIT Press, 1992); David Papineau, "Representation and Explanation," *Philosophy of Science* 51, 4 (December, 1984), pp. 550-572; David Papineau, *Reality and Representation* (Oxford UK, Cambridge USA: Blackwell, 1987); David Papineau, *Philosophical Naturalism* (Oxford UK, Cambridge USA: Blackwell, 1993); Ruth Millikan, "Biosemantics," *Journal of Philosophy* 86, no. 6 (June 1989), pp. 281-297; Ruth Millikan, *Language, Thought and Other Biological Categories: New foundations for realism* (Cambridge, Mass.; London, England: MIT Press, 1984); Robert Cummins, *Meaning and Mental Representation* (Cambridge, London: MIT Press, 1989).

school, as a poor student, I will trade having a full stomach with having the opportunity for enlightenment. I can really think about many things, things that are present in my room now, things that exist somewhere but not in my room, things in the future, things that happened in the past, and also things that have never existed and will never exist.⁶ My mind is, most of the time, directed towards or about something in the world. In Brentano's terminology, this is called "intentionality", and it is, according to Brentano, what draws the sharp line between the mental and the material.⁷

One of the effective ways to describe intentionality is that intentional systems differ from non-intentional systems in that states of the former have the capacity to represent things in the world as being in a certain way, whereas states of the latter do not have any capacity to represent anything at all. At this point, the central issue of intentionality appears to be based on the very idea of representation: "[H]ow can a mere mechanism think about and represent things?"⁸

There have been two different, but most of the time related, positions in the history of philosophy concerning the very idea of mental representation.⁹ The first position accepts the obvious fact that part of being a subject is to be able to represent the world. No one in the history

⁶ For a nice introduction to intentionality, see Hugh Clapin, "Introduction," pp. 1-2; Tim Crane, *The Mechanical Mind*, pp. 30-41. For an introduction to its historical background, see Tim Crane, "Intentionality." In *Routledge Encyclopedia of Philosophy*, edited by Edward Craig (London, New York: Routledge, 1996), vol. 4.

⁷ Most people now disagree with Brentano's thesis that the mental and material are distinguished by intentionality. Nevertheless, they all say that intentionality is important and it should be given an account within a naturalist framework. For a discussion of Brentano's thesis in the context of contemporary philosophy of mind, see Tim Crane, *The Mechanical Mind*, pp. 36-40.

⁸ Tim Crane, *The Mechanical Mind*, p. 6.

⁹ I owe the following point to Stephen Voss.

of philosophy, including the archenemies of the idea of “mental representation” such as Ryle and Wittgenstein, probably rejects such an approach to explain what it is to be a subject. If I am a subject and believe that it is snowing now, then I am in a certain state of mind that represents the world as being a certain way. Consider other instances of such representational states of different subjects. Zeynep hopes that it is snowing now, Hatice desires that it is snowing now, Salih fears that it is snowing now, Özgür doubts that it is snowing now and Betül regrets that it is snowing now. These states are states of different subjects and each of these states stand for a different intentional/representational attitude one takes towards the same mental content, the content expressed by the sentence “it is snowing now”.¹⁰ This content represents a certain state of affairs outside, that is, the state of snowing now. It is obvious that different subjects can have the same attitude toward it, and the same subject can have different attitudes toward it.¹¹ Furthermore, it appears that what makes the difference between two particular beliefs must have to do with their different contents. I have a belief that it is snowing now, and this belief is different from my other belief that there is an ice cream in the fridge. So, different beliefs are individuated by their different

¹⁰ These examples might give the impression that all intentional/representational states are propositional attitudes. This is not the case. There can be non-propositional attitudes which yet represent something. For instance, my nephew Salih, as a baby, feared vacuum cleaners. He had an intentional/representational state towards the vacuum cleaners, but he did not presumably have a propositional attitude. So, the basic idea here is that thoughts, which include diverse sorts of mental states (love, fear, belief, desire, hope, etc.), are all representational. You just cannot love without loving something/somebody, and by loving something/somebody you represent the individual object as being a certain way. I follow Crane’s view here. See *Tim Crane, The Mechanical Mind*, pp. 25-26.

¹¹ Jaegwon Kim, *Philosophy of Mind* (Boulder, Col.: Westview Press, 1996), p. 184.

contents.¹² The same goes for other sorts of mental states. Of course, this way of describing mental beings does not compel one to accept a further claim that there exist mental representations in subjects' minds, mental structures that stand for states of affairs, objects and relations in the world. So, the first position only states that a subject can entertain a representational state without having mental representations in his mind. The second position, however, insists on the fact that we have (authentic) intentional/representational states due to the presence/storage and manipulation of sensory and cognitive representations in our mind. In virtue of those mental representations a subject can have intentional states about the world. In other words, intentional states owe their intentional characteristics to what constitute them, that is, mental representations.¹³ That is the one and only way, from the perspective of this approach, for a subject to relate himself to the world via his mental representations and fulfill his desires provided that the world portrayed by his mental representations is real and accurate, and also the real world is cooperative with his "rational" desires.

The Problem of Mental Representation

In the previous section, I presented the difference between two different but related positions in the history of philosophy. Furthermore, I give a short description of intentionality. If one holds the second position, that

¹² Crane, *The Mechanical Mind*, p. 25.

¹³ Robert A. Wilson, "Philosophy." In *The MIT Encyclopedia of Cognitive Sciences*, edited by Robert A. Wilson and Frank C. Keil (Cambridge Mass.; London, England: The MIT Press, 1999), p. xxvii.

is, the claim that mental representations are essential ingredients for a being to entertain intentional / representational states, then he / she is forced to answer one of the big questions in the history of philosophy. In a notorious letter to Marcus Herz, Kant gives us probably the finest description of the problem:

As I thought through the theoretical part, considering its whole scope and the reciprocal relations of all its parts, I noticed that I still lacked something essential, something that in my long metaphysical studies I, as well as others, had failed to pay attention to and that, in fact, constitutes the key to the whole secret of hitherto still obscure metaphysics. I asked myself: What is the ground of the relation of that in us which we call "representation" to the object?¹⁴

Kant's problem can be formulated in a more general form: What is it for anything (word, picture, building, music notation, number, thermostat, data structures in a computer, rings of a tree, dark clouds, facial expressions, dance movements in syrtaki, dance of a honeybee, state of mind, and so forth) to represent anything else?¹⁵ As for conventional representations, ones that inherit their representational power from the minds of users and conventions of different societies of minds, the answer is straightforwardly and relatively easier than a possible answer, for instance, to explain natural representations and mental

¹⁴ Immanuel Kant, *Kant: Philosophical Correspondence, 1759-99*, edited and translated by Arnulf Zweig (Chicago: Chicago University Press, 1967), pp. 70-75. I will not get into the details of how this question, within the context of Kant's philosophy at that time, poses such a big threat for Kant. For a very nice discussion of this point, see Bülent Gözkan, "Çevirenin Sunuşu." In Gottlob Frege, *Aritmetiğin Temelleri*, translated by Bülent Gözkan (Istanbul: Yapı Kredi Yayınları, 2008), pp. 25-31. For a discussion of Kant's problem within the context of contemporary philosophy of mind, see Hilary Putnam, "Does Evolution Explain Representation?" In *Renewing Philosophy* (Cambridge, London: Harvard University Press, 1992), pp. 19-34.

¹⁵ Tim Crane, *The Mechanical Mind*, p. 11. See also Robert Cummins, *Meaning and Mental Representation*; Nelson Goodman, *Languages of Art: An approach to a theory of symbols* (London: Oxford University Press, 1969).

representations. Words, pictures, etc., “need interpreting... interpretation is something which the mind bestows upon words. Words and pictures gain the interpretations they do, and therefore represent what they do, because of the states of mind of those who use them”.¹⁶ Putnam also emphasizes exactly this point. For Putnam too, lines in the sand, noises, things that occur as a result of intentions, conventions of the users do not have the power, in themselves, to represent anything.¹⁷

This way of thinking has led some philosophers, including Tim Crane, to hold the view that all the rest of representations, that is to say, natural and conventional representations, depend on the presence of minds.¹⁸ Without the interpretative activities of mental beings, there would not be kinds of representations such as natural and conventional representations. Hence, mental representations are prior to the rest of representations, and we must put all our effort into understanding how the mind, in itself, can represent anything. “[H]ow is it that thought forms can ‘in themselves’ represent anything? Or can they? How can thought reach out and ‘grasp’ what is external?”¹⁹

A group of philosophers, on the other hand, objects to the part of the claim above that mental representations are prior to natural representations, although they agree with the other part of the claim that mental representations are prior to conventional representations.

¹⁶ Crane, *The Mechanical Mind*, p. 22.

¹⁷ Hilary Putnam, *Reason, Truth and History*, p. 2.

¹⁸ See Crane, *The Mechanical Mind*, p. 13. See also Gerard O’Brien and Jon Opie, “Towards a Structural Theory of Mental Representation.” In *Representation in Mind: New Approaches to Mental Representation*, edited by Hugh Clapin, Phillip Staines, Peter Slezak (Amsterdam; Boston: Elsevier, 2004), pp. 4-5.

¹⁹ Putnam, *Reason, Truth and History*, p. 2.

Following Grice's footsteps in his seminal work "Meaning",²⁰ these philosophers attempted to build and develop the view that mental representations are actually a species of natural representations. So, they believe, if we can find a way to understand natural representations, we will be able to shed some light on the nature of mental representations as well. Dretske is one of the eminent figures of such an approach.

In the end, both groups seek to find an answer for the same question, albeit by following different approaches. In this thesis, I will also attempt to find an answer for the same question. The question about the nature of mental representation then is the following one: what is the ground of mental representations?

Before I start presenting the topic of the subsequent chapters of this thesis work, I would like to introduce the problem of error, which is regarded as the biggest issue especially in the contemporary literature on mental representations. This problem can be briefly expressed in the following way: Given the obvious fact that it is most likely that there can be a gap between the way a subject of mind represents the world and the way the world is, then we must accept that some of our representations may turn out to be false. If a theory of mental representation cannot give an account for how this can happen, then it won't be satisfying the most important requirement: no possibility of error, no representing either.²¹

²⁰ Paul Grice, "Meaning," *Philosophical Review* 66, no. 3 (Jul., 1957), pp. 377-388.

²¹ Crane, *The Mechanical Mind*, pp. 175-177; Neander, Karen. Spring 2012. Teleological Theories of Mental Content. Available [online]: "<http://plato.stanford.edu/archives/spr2012/entries/content-teleological/> [Spring 2012]".

In the contemporary literature, the problem of error is generally explained in two different, but related forms. I will call the first one “the problem of misrepresentation”, and the second one “the disjunction problem”. Let’s go over them one by one.

Let’s suppose the following definition of representation to be true: X represents Y iff Y is the cause of X. So, for instance, we can say, on this account, that my dog-image or my dog-thought represents dog because dog is the cause of my mental representation. Let’s now consider whether this definition of representation can give an account for the cases of misrepresentation, the cases in which the represented object is not present or the represented object is not represented correctly.²² For instance, can there be a dog-image in my mind without the presence of a dog in my environment? The definition above does not allow any misrepresentation to occur, because X represents only whatever causes it. So, if there is not a dog in my environment, and if something else, say wolf, is present, then I should be representing a wolf. No gap can occur between the way I represent the world and the way the world as it is. So, the account in question simply rules out the cases in which the represented object is not present or the represented object is not

²² Let’s take the following formulation as a particular form of the general format above: A represents B iff B is the cause of A. Here, A stands for “dog-image” or “dog-thought” in my mind, and A will be representing a dog. We can then explain the cases of misrepresentation in two ways. Firstly, since the represented object here is DOG and dog is not present in my environment to token a mental representation DOG in my mind, and if a wolf in my environment is actually responsible for tokening a mental representation DOG, then I will be misrepresenting the world. Secondly, A may stand for a “black-dog-image” or “black-dog-thought” in my mind, but the cause (B) of that mental representation might be a white dog. So in this case, we say that when the represented object DOG is actually present in my environment but the way I represent it (for instance, a black-dog-image) is not true of that dog (a white dog) causing my mental representation, we have a case of misrepresentation. Of course, these are cases of misrepresentation in sensation. There are also cases of misrepresentation in thinking. In this thesis, I will be focusing mostly on cases of misrepresentation in sensation.

represented correctly.²³ Therefore, it will never be possible for X to misrepresent. So, error is not possible.

In the literature, it is often thought that the problem of misrepresentation has a close cousin, the so called “the disjunction problem”.²⁴ Suppose that I have a mental representation of a dog when there is a dog around me. But, it is possible to have a mental representation of a dog when, for instance, the cause of that representation is a wolf or a decoy dog. Let’s say that there is a systematic correlation between my mental representation DOG and the causes of DOG. Then, I should say that whenever I have a token of mental representation DOG, DOG should represent whatever caused my DOG-representation to occur. DOG then would have a disjunctive content, for instance, “either a dog or a wolf is present”. Were my representations to have disjunctive content, it would fill the entire gap between the way the world as it is and the way I represent it to be. Then, error again is not possible. In this thesis work, I will discuss whether theories of representation presented can deal with the problem of error.

Précis of chapters

Aristotle’s theory of mind is probably one of the earliest sources for the idea of mental representation. In Chapter 2, I will first present and critically assess Aristotle’s account of sensation and thinking in *De Anima*. As far as I understand, Aristotle’s account of mental representation is

²³ Cf. Crane, *The Mechanical Mind*, p. 181.

²⁴ Ibid., p. 179.

grounded in two factors: (1) causation and (2) (some kind of) resemblance. In the end of the chapter, I will also talk about the way Aristotle explains the source of misrepresentation.

Of course, so much has been changed in the twentieth century philosophy of mind since Aristotle. So, in order to understand the mainstream literature on mental representation, especially between the 1980s and 1990s, one needs to be clear about the underlying metaphysical framework at the background, which is, most often, a functionalist metaphysics of mind. However, it is hard to describe functionalism without making a comparison to previous popular metaphysics such as behaviorism and type physicalism before the debut of functionalism. For this reason, in Chapter 3, I evaluate and critically examine three different theories of mind (behaviorism, type physicalism and functionalism) in the twentieth century philosophy of mind, each of which offers radically different accounts for the metaphysical questions like the following:

“what is it for a subject to represent something else?” “In general, what is it to have a mind / mental states?” For behaviorism, mind is, in essence, a behavioral kind. Type physicalism, rejecting the acausal account of behaviorism, explains mind in terms of the causal power of neural states mediating between a set of stimuli and behavioral responses.

Functionalism, on the other hand, rejects type-identifying mental states with neural states and offers a very liberal account of mind, according to which mind is identified as a functional kind (of appropriate complexity).

In the last section of this chapter, I maintain the idea that functionalism actually rests on a very Platonic ideology, and I attempt to show how this Platonic ideology manifests itself through the most popular example of

functionalist thought, that is, the software/hardware distinction. Then, I explain how such a Platonic reading of functionalism can be very problematic, and, I suggest that the Aristotelian way of interpreting functionalism does seem to be a better philosophical venue for a common-sensical theory of mind. According to this interpretation, mind can be housed only within a limited range of physical/spiritual structures, and the function and matter (physical or spiritual stuff) of mental states, contrary to the often-applied Platonic tendencies one may have, are very much connected. In short, I believe that mind is, at bottom, a functional kind (in the Aristotelian sense). This is the conclusion I have arrived, at the end of the metaphysical quest for the nature of mind in the third chapter.

The target of Chapter 3 is not only to introduce functionalism as a new metaphysical framework about the nature of mind. It is also meant to prepare the groundwork for the later chapters. In particular, it is very crucial to understand what functionalism is, because, in one way or other, most philosophers have become functionalists since Putnam's essay "Psychological Predicates". So, in order to understand Dretske and Millikan's theories of mental representation in later chapters, one needs to know beforehand the very essence of functionalism.

After having introduced and discussed functionalism, in Chapter 4 I will talk about the transition from early functionalist programmes (Putnam) to post-functionalism programmes such as Dretske's indication theory and Millikan's teleosemantic theory. Functionalism was not only a new metaphysics of mind in the twentieth century philosophy of mind, it was also meant to give an account for all kinds of

intentional/representational states. The crucial point in Putnam's essay "Psychological Predicates" was that functionalism was supposed to explain all kinds of intentional/representational states, ranging from relatively simple states like pain to complex states like the thought about the next summer's vacation. So, for instance, pain is identified in the following manner:

"Being in pain = being the fifth of n states, $S_1, \dots, S_n$, whose relations to one another and to inputs and outputs are specified by FO ($S_1, \dots, S_n, i_1, \dots, i_k, o_1, \dots, o_l$)."²⁵

Likewise, being in a state of thinking about next summer's vacation then would be described by the same formula, except that "being in state S_5 " is replaced with, for instance, "being in state S_{107} " along with input and output clauses.²⁶

After writing "The Meaning of 'Meaning'",²⁷ however, Putnam realized that the very ideology of functionalism is quite utopian and science-fictional. Of course, people who are sympathetic to original functionalism did not immediately abandon functionalism. Among those post-functionalist philosophers come Dretske and Millikan.

²⁵ Oron Shagrir, "The rise and fall of computational functionalism," In *Hilary Putnam (Contemporary Philosophy in Focus)*, edited by Yemima Ben-Menahem (Cambridge, New York: Cambridge University Press, 2005), p. 230. FO stands for "functional organisation".

²⁶ Ibid., p. 230.

²⁷ Hilary Putnam, "The Meaning of 'Meaning.'" In *Language, Mind and Knowledge*, edited by K. Gunderson. Minnesota Studies in the Philosophy of Science, Vol. 7. (Minneapolis: University of Minnesota Press, 1975), pp. 131–193. Reprinted in Hilary Putnam, *Mind, Language and Reality, Philosophical Papers, Volume 2*. (Cambridge: Cambridge University Press, 1975), pp. 215–271.

In Chapter 5, I will present Dretske's theory of representation in his pioneering work *Knowledge and the Flow of Information*.²⁸ According to Dretske's view, mental representations are species of natural representations. So, understanding how natural representations represent will be the initial step towards understanding the nature of mental representations. That's the reason why Dretske begins with Grice's idea of "natural meaning", and develops this idea in order to offer a philosophical theory of information/representation, according to which information/representation is grounded in law-like correlations between the source and receiver. Dark clouds represent rain because of law-like correlations between dark clouds and rain. Smoke represents fire because of law-like correlations between smoke and fire. Likewise, representing vehicles in the mind represent something as F because of the law-like correlations between those representing vehicles and the objects with F. After introducing Dretske's view, in the end of chapter, I will deal with the question whether Dretske's theory can solve the problem of misrepresentation.

In Chapter 6, I will present Millikan's theory of (biological) representation in her famous article "Biosemantics".²⁹ First, I will explain Millikan's criticism of Dretske's theory of representation. Then, I will begin introducing Millikan's way of explaining mental representation. The most important concept for Millikan is "proper function". A defective heart, for instance, may not perform its proper function, but we still say that pumping blood is its proper function. On the other hand, a

²⁸ Fred I. Dretske, *Knowledge and The Flow Of Information* (Stanford, CA: CSLI Publications, 1999; originally published: Cambridge, Mass.: MIT Press, c 1981).

²⁹ Ruth Millikan, "Biosemantics," *Journal of Philosophy* 86, no. 6 (June 1989), pp. 281-297.

molecular duplicate of me that emerges out of cosmic accident, like in the case of Davidson's swampman, can have mechanisms that are identical with the mechanisms of a biological system, but, Millikan believes, the molecular duplicate would not have the right history, and therefore, its mechanisms (a heart, a kidney, an eye or a brain) would not have a proper function. If something does not have a proper function, then it is not a biological category. This is very crucial for Millikan, because she thinks that mentality, at bottom, is a biological phenomenon. In the end, I will deal with the question whether Millikan's theory can explain the possibility of misrepresentation. It does seem to me that, once you agree with the basics of Millikan's theory, misrepresentation is no longer a problem.

In Chapter 7, I will wrap up the whole discussion in this thesis within the context of the problem of mental representation. Having provided enough material about the journey of the idea of mental representation in the history of philosophy, I will discuss the virtues and vices of the theories of mental representation presented in the thesis. As far as I know, contrary to Fodor's conviction, there have been three main options in the history of philosophy that are proposed as the ground of mental representation, which are:

(R1) X represents Y iff there is (some sort of) resemblance between X and Y

(R2) X represents Y iff there is a causal relationship between X and Y

(R3) X represents Y iff (i) there is a causal relationship between X and Y, and (ii) there is (some sort of) resemblance between X and Y

Millikan's teleosemantic theory of mental representation seems to be a better theory than Dretske's information theory of representation. However, this is not to say that it is a trouble-free theory of mind. In the second section of this chapter, I will talk about the trouble with the contemporary literature on theories of mental representation in general and teleosemantics in particular. One of the working hypotheses behind the research strategies for probably most of the theories of mental representation in the late twentieth century was that any theory of mental representation must begin with the analysis of simple mental states/capacities and if the theory in question achieves a considerable success in explaining those simple ones, it will deal with complex mental states and capacities later.

As an offspring of this working hypothesis, most of the debates in this era revolve around the questions about fixing the content of simple representational states. As far as I can see, the literature got stuck with the first step (explaining simple mental states/capacities) and has not even managed to move to the next step (explaining complex mental states/capacities). A quick look at the debates about the frog's representation of a fly is only one instance indicating the trouble with the literature. I believe it is now just the right time to take the next task and start considering the questions like, for instance, whether there can be teleosemantic explanation for the complex representational states, for instance, "The Arab Spring has been a quite influential movement in the Middle East". It seems to me that cases of this sort are outside the territory of teleosemantic explanation. So, while it may have been true of teleosemantic explanation that it can give an account of simple mental

states, the same does not hold to be true for complex mental states. I don't see any way to develop teleosemantic theory to give an account of complex mental states, and therefore I abandon it.

In the last section of Chapter 7, I will revisit Aristotle's account. Aristotle's account differs from the post-functionalist accounts defended by Dretske and Millikan, because, unlike Dretske and Millikan, it is not built on a Platonic metaphysics. Furthermore, there is a very crucial point in Aristotle, which, I believe, indicates exactly the missing perspective in the views of Dretske and Millikan. In this section, first I will summarize some of the main points in Aristotle's account of mind as a representing capacity, and then I will focus on his distinction between philosophical definition and physical/natural definition. I believe that understanding this difference helps us understand the good direction towards understanding the nature of mental representation, and show the source of the chronic problem within post-functionalist naturalist attempts to understand mind and representation. Following Aristotle, I believe that the fact that mind is part of nature does not mean that representation, as an essential aspect of mind, consists in a physical/material process. In understanding the nature of mind as a representing capacity, we should also look for the formal criterion for representation as Aristotle does in *De Anima*. The philosophical investigation should not only limit itself to the physical/material explanation, it should go further and give a formal/conceptual analysis of what is being represented and misrepresented in mind. This is very demanding and it really pushes one to practice philosophy as excellently as possible. Following Aristotle's rigorous practice of philosophy, I think we should take the same attitude

when we deal with the very nature of mind as a representing capacity. Representation is a very complex intellectual capacity, and we just cannot explain it through a materialist functionalist method, the method that is employed by Dretske and Millikan. If you take representation in isolation from other mental states and try to give a materialist/ functionalist description of representation, you will definitely end up with a sort of difficulty in explaining the ridiculous cases such as air representing wind, water representing the temperature, and you will not ever come close to understanding the very nature of mental representation. These sorts of causal theories always forget to mention the formal aspect of representation as in the case of a physicist in *De Anima* who forgets to mention that anger is partly constituted by a desire to harm. Likewise, representing is partly constituted by, among other things, desiring to find a way back home, hoping to understand what other human beings represent, abstracting the object of thought from its particular features in a space-time continuum, associating different sorts of objects under a new class, writing a story about one's experience of the world. Representation, as Aristotle would say, is essentially/ conceptually related to both perceiving and thinking. That is the nature of representation, which very much resists any sort of material reduction.

CHAPTER II

ARISTOTLE'S THEORY OF SOUL AND MIND³⁰

"The body's got a soul too, too, have pity on it..."³¹

Aristotle's theory of soul is arguably one of the most interesting accounts in the history of philosophy that have been offered to explain the nature of psychological capacities. First of all, his account, in opposition to Plato, seems to support the idea that sensation is a species of thought. Secondly, it was probably the first fine instance of a theory of mental representation. Thirdly, by introducing the notion of representation, it particularly emphasized the central role mental representations play in one's faculty of imagination and thought. In this respect, Aristotle's theory of soul was most probably the birthplace of the concept of representation in the history of philosophy.³² In this chapter, I will primarily aim to present and critically evaluate Aristotle's theory of soul in *De Anima*³³ with a special focus on his account of sensation and thought. In the first section, I will present a brief outline of his understanding of soul. In the second section, I will explain conceptual

³⁰ I owe special thanks to Chryssi Sidiropoulou and all the members of her *De Anima* class (Fall2012) for their comments, critical questions and suggestions.

³¹ Nikos Kazantzakis, *Zorba the Greek*, translated by Carl Wildman (London: Faber and Faber, 1961), p. 36.

³² Aristotle's theory of representation has been developed, modified and criticized in the medieval ages. For an exhaustive evaluation of Aristotle's impact on medieval age thinkers, see Henrik Lagerlund. Fall 2008. Mental Representation in Medieval Philosophy. Available [online]=
"<http://plato.stanford.edu/archives/fall2008/entries/representation-medieval/>" [Fall 2008].

³³ Aristotle, *On The Soul; Parva Naturalia; On Breath*, translated by W. S. Hett (Cambridge, Mass.: Harvard University Press, 2000, c1957.) I will make use of Hett's translation, but, as it is commonplace in the literature, I will refer to Aristotle's work as *De Anima* and cite the work by Bekker page and line numbers.

details hidden behind Aristotle's approach to perception. In the third section, I will focus on the connection between imagination and mind, and I will talk about his distinction between agent mind and patient mind.³⁴ I will finish with some remarks on a possible Aristotelian way to solve the problem of error.

Introduction to Aristotle's Soul

In our present time, psychology is often understood as a discipline investigating, with its own research methods, various features of soul in animals and human beings such as perception, memory, emotion and so forth, and their relations with the environmental and social parameters/factors, and their interrelations among each other. Compared to this conception of soul, it is very obvious that Aristotle understands a radically different sort of conception from the word "soul".³⁵ In Aristotle's terminology, the soul is conceived as the principle of life, something that makes beings alive and its separation makes those beings dead. In a technical vocabulary of Aristotle's metaphysics, soul is an essential form of the bodies of individual beings definitive of their characters in this world. For instance, the soul of human bodies is constituted by nutritive, sensitive and cognitive powers defining the essential features of being a human (i.e., rational animal), whereas the

³⁴ There is almost a common trend to refer to Aristotle's *De Anima* as the source of the naive view of the similarity theory. Contrary to this common trend, I found Putnam's remark quite helpful in finding the right direction to interpret his hybrid theory. See Hilary Putnam, *Reason, Truth and History* (Cambridge, New York: Cambridge University Press, 1981), pp. 57-58.

³⁵ For an exhaustive discussion about the background impact of ancient Greek context and Plato on Aristotle's thoughts, see Daniel N. Robinson, *Aristotle's Psychology* (S. I]: Joe Christensen Inc.: 1999), Chapter 1-2.

soul of plant bodies is constituted only by nutritive power of the soul defining the substantial aspect of being a plant. Accordingly then, from the term “soul” Aristotle understands a very broad class of entities revealing psychic activities, some of which (e.g., plants) would be classified as soulless by the contemporary psychology.

To begin with, Aristotle considers soul as an essential form of the bodies without which those bodies encapsulating soul would be regarded as lifeless. In this picture, soul is very much connected to the body; in a sense as an energy activating the potential states of the body and helps that body show life signs.³⁶ Again in this picture, soul is divided into three main parts as nutritive, sensitive and cognitive powers with respect to its different life activities; and there is an order of hierarchy among those powers of the soul:

“The first principle in plants, too, seems to be a kind of soul; for this principle alone is common to both animals and plants. It *can* exist in separation from the sensitive principle, but nothing *can* have sensation *without* it.”³⁷

It is quite obvious and significant, for him, to notice the natural demarcation among living beings, according to which some of the living beings only partake the nutritive power of soul, whereas others have both

³⁶ Here and some other places in this section, I made use of a metaphor “energy flowing through a bodily channel” in order to help the reader what Aristotle understands by his statement that soul is the form of body. But, I never intend to raise a suspicion, by the use of such metaphors, that Aristotle, very much like what Plato would understand, thinks of the soul as a miserable, unhappy prisoner in the body with forgotten memories. I wanted to make it clear, at least in a footnote, that Aristotle is wholly against such a conception of soul, and one must understand that soul, in Aristotle’s metaphysics, is an important part of nature since human beings qua living beings are part of this big system of nature. So, nowhere in this chapter do I want to picture Aristotle’s notion of soul in a spiritualist manner.

³⁷ Aristotle, *De Anima*, 411b28-30. Italics are mine.

nutritive and sensitive powers of soul and, moreover, some other sophisticated beings entertain cognitive/intellectual power of the soul in addition to those aforementioned ones.³⁸ This natural demarcation implies a natural order of hierarchy among the powers of soul beginning with the lowest degree of soul, i.e., nutrition, and finishing with the highest degree of soul, i.e., thinking and judging. Consequently, it eventually seems naturally necessary for a living being with a mind, like a human being, to have sensitive and nutritive powers; and for a living being with sensation to have nutritive power as well. It is important that the reverse order of hierarchy does not hold to be true.³⁹ Put differently, living beings standing in the middle stage or highest stage of this order of hierarchy contain the powers of soul in the lower stages, whereas living beings standing either in the lowest stage or the middle stage will not get any power of soul in the higher parts.

The natural hierarchy among the powers of soul is an interesting aspect of Aristotle's theory and it seems to be in complete harmony with his definition of soul. Indeed, one can plausibly say that Aristotle's theory of soul, in general lines, is very much committed to the idea that soul is form of body and related to it in an essential way as the Hermes-figure/form is related to the bronze material in constituting an Hermes statue. In accordance with this approach plus Aristotle's doctrine on the natural hierarchy among operations of the soul, one would expect from

³⁸ Ibid., 413b0-15; 414b0-20. Also see especially 434a22-434b9 in order to examine Aristotle's remark about the natural demarcation among the powers of soul and his reasoning about the natural connection between sensation and nutrition.

³⁹ Shields, Christopher. Winter 2008. Aristotle's Psychology. Available [online]= "<http://plato.stanford.edu/archives/win2008/entries/aristotle-psychology/>" [Winter 2008].

Aristotle that powers of the soul be always related to the body in an inseparable way-except for death. However, he is especially very suspicious about the status of one power of the soul, i.e., intellect. It seems to follow as a generalization from the natural hierarchy that living beings with minds must also be individual entities with sensation and nutrition. But, he observes the fact that while it is possible to associate other powers of the soul such as nutritive and sensitive parts with respectively corresponding features of the body such as nutrition organs and sense organs, it is not that possible to find a bodily channel through which the mind realizes its power.⁴⁰ Aristotle really seems to be deeply puzzled about the intractable nature of mind.

It seems, for many people, that there is a very sharp tension between his claim that soul is the form of the body of an individual being and his other claim that mind, as a power of the soul, is separable from the body.⁴¹ As a consequence of this tension, one may tentatively argue that it would be no good reason to conclude, as a natural implication of the hierarchy among the powers of soul, that nature dictates us to think that a being with a mind must have sensitive and nutritive faculties as well, and therefore that being must always be a material entity at the

⁴⁰ I will open up this issue later on in this chapter.

⁴¹ It is not very certain whether Aristotle is definite whether mind is separable from the body. One can find some passages in *De Anima* where Aristotle seems to be inclined to think of mind as independent of the body. But, one can also find some passages where Aristotle is against this claim. For instance, if you take his claim “no organ for the mind” as a very strong claim, you may say that this must lead to the corollary that mind can persist through time and it can exist without the body. But, on the other hand, given Aristotle’s commitment to the idea that all forms of thinking require imagination, and imagination requires the body, you may be inclined to say that even though Aristotle explicitly denies any bodily organ for the mind, this does not necessarily lead to the corollary above. For instance, you can say that thinking requires the body. I owe this point to Stephen Voss.

same time. It could be stated that Aristotle's theory of soul is generally a materialist and functionalist sort in regard to its explanatory method under which the nature of states of soul is explicated in terms of their functions in a material entity; yet it is very controversial to hold the materialist claim about his approach to the mind,⁴² given the fact that Aristotle seems to be seriously considering the possibility of a being with a mind as separable from the body and that this possibility presumably requires a certain way of modification to be made in the hierarchy of the soul. Let this suffice for the general outline of Aristotle's theory of soul, at least for the purposes of this chapter. It will be more helpful and useful now, after drawing the basic lines of his theory, to concentrate on his account of sensation and thinking, and show the soul of the similarity theory of representation implicit in that account.

Sensation

Before evaluating Aristotle's account of sensation, it is crucial to briefly step into his understanding of change. There are two kinds of change, the first one that involves the change of the essential properties, either the form or the matter, of an individual being destroying its identity as that individual being, e.g., destruction of the Hermes-form of an Hermes statue, destruction of the bronze material of an Hermes statue; and the second one that involves change of accidental properties of an individual

⁴² Shields, *Aristotle's Psychology*; Deborah K. W. Modrak, "The Nous-Body Problem in Aristotle," *Review of Metaphysics* 44, no. 4 (June 1991), pp. 755-774; Charles H. Kahn. "Aristotle on Thinking." In *Essays on Aristotle's De Anima*, edited by Martha C. Nussbaum and Amelie Oksenberg Rorty (Oxford [England]: Clarendon Press, 1997), pp. 359-380.

being not resulting in the loss of its identity as that individual being, e.g., painting a house a white color. Aristotle applies to the second conception of change in the explanation of perceptual occurrences of soul.⁴³ This must be kept in mind in order to understand Aristotle's conception of sensation.

"Thus during the process of being acted upon it is unlike, but at the end of the process it has become like that object, and shares its quality."⁴⁴

"[S]ense is that which is receptive of the form of sensible objects without the matter..."⁴⁵

For any kind of change to happen—perhaps besides internal change of states in subjects—, there must be two objects. Let us call the one that is acting upon the other object agent object and the other one that is receiving that effect the patient object. In fact, Aristotle thinks that the same sort of thing is going on while we perceive an object. In perceptual occurrences of the soul, there is typically an object outside acting upon one of our receptive capacities and this object gradually changes the sensible quality of the receptive capacities. In this framework, there must be, of course, a line drawing the difference between non-perceptual and

⁴³ Aristotle, *De Anima*, 416b33-417a2; Shields, *Aristotle's Psychology*.

⁴⁴ Aristotle, *De Anima*, 418a3-6.

⁴⁵ Ibid., 424a16-20. There is a very interesting scholarly controversy whether Aristotle's theory of soul is a functionalist sort or not, and this sentence is a kind of starting point for the debate on this matter. For this controversy, see M. F. Burnyeat. "Is an Aristotelian Philosophy of Mind Still Credible? (A Draft)." In *Essays on Aristotle's De Anima*, edited by Martha C. Nussbaum and Amelie Oksenberg Rorty (Oxford [England]: Clarendon Press, 1997); and Martha C. Nussbaum and Hilary Putnam. "Changing Aristotle's Mind." In *Essays on Aristotle's De Anima*, edited by Martha C. Nussbaum and Amelie Oksenberg Rorty (Oxford [England]: Clarendon Press, 1997). Also see Christopher, Shields. "The First Functionalist." In *Historical Foundations of Cognitive Science*, edited by J. C. Smith (Dordrecht: Kluwer Academic Publishers, 1991), pp. 19-34; Christopher D. Green, "The Thoroughly Modern Aristotle: Was Aristotle really a functionalist?" *History of Psychology* 1, no. 8 (February 1998), pp. 8-20.

perceptual changes.⁴⁶ The mere occurrence of a sensible quality in a patient emerging as an effect of the agent is not enough to describe that patient as perceiving something; because one can think of the case of garlic smell impressed in the air without attributing any perceptual state to air. So, Aristotle claims that the agent must be one capable of transferring its sensible qualities and the patient must have a sense organ that in which capacity to receive sensible forms resides,⁴⁷ and besides there must be a real object outside (not imaginary object or object of hallucination) as an agent. In addition to all of this, there must be a

⁴⁶ Aristotle, *De Anima*, 424a17-424b20. This is one of the passages in *De Anima* that requires utmost attention; unfortunately the passage is very short and Aristotle does not give enough detail of what he has in mind. We get a very good reconstruction of this passage in Aquinas' commentary. The problem can be formulated in what follows: if we describe sensation over the patient object's capacity to receive sensible forms, then we cannot differentiate the case of air receiving sensible forms from the case of animal receiving sensible forms since it is obvious that both have capacity to receive sensible forms, and therefore we need to accept that objects like air, light also perceive sensible forms of objects. Aquinas launches his understanding of the passage by introducing a distinction between material mode of existence and cognitive/spiritual mode of existence for forms. According to Aquinas, the sensible form received by air indicates a material mode of existence where form is not completely separated from matter, and the sensible form received by animal indicates a cognitive/spiritual mode of existence where this form is completely separated from matter, and this is the reason why Aristotle claims that senses receive forms without matter. See St. Thomas Aquinas, *Commentary on Aristotle's De Anima*, translated by Kenelm Foster and Silvester Humphries (Notre Dame, Indiana: Dumb Ox Books, c1994), pp. 171-174.

⁴⁷ Aristotle, *De Anima*, 424a20-30. For Aquinas, there is a very specific reason behind Aristotle's introduction of a sense organ in the middle of the discussion about perceptual and non-perceptual changes. People may read the claim that sense receives forms into cognition immaterially and take Aristotle to have the goal that sense, like intellect, must be an immaterial power. However, Aquinas suggests, Aristotle is very firm on his idea that power of sensation is realized by sense organs. See Aquinas, pp. 172-173. Although I agree with Aquinas on this interpretation, I believe that there is another agenda in this passage. After the statement where Aristotle mentions sense organ as that in which a power of sensation resides, he deals with an example of violent effect of sense object which a sense organ is naturally unable to bear. Of course, he does not compare this case with the case of air receiving sensible forms. But, I think that it is plausible to interpret him as saying that no matter how violent is the effect of sense-object on air, air will not have any damage whatsoever. However, the violent impact of the sense object in the case of sense organ could be very painful and result in temporary or permanent loss of that power. This can be considered to show that sensation is dependent on a sense organ that is limited in certain respects in accordance with the function of that being using that sense organ, but air does not have a sense organ and cannot be said to be in a state of sensation. In order to have sensation, a patient object must have, among other things, a sense organ to navigate his environment and help that being to utilize its purpose on earth.

medium of sensation presenting sensible forms before the sense organ of the patient object.⁴⁸ Only after all these general conditions are satisfied it is reasonable, from Aristotle's perspective, to describe the patient as perceiving something. Then, by following Aristotle's receipt, we have something like the formula below for the explanation of perception:

A subject S perceives an object, X iff (1) X is present in the perceptual field / domain of that subject,⁴⁹ (2) S has a sense-organ capable of receiving the sensible forms of X, (3) X acts upon that relevant sense-organ by transferring / imprinting its sensible forms / qualities to it through a medium of perception,⁵⁰ and (4) that relevant sense-organ, as a consequence of interaction with X, resembles X with respect to the sensible qualities / forms of the object in action (that sense-organ becomes isomorphic with the sensible forms of X).

Now, let us consider the following example in order to grasp what this formula tells us. In the morning, as I enjoy regularly each day, I drank a cup of tea; only this time my pleasure of having a cup of tea turned out to be a real pain. I was not careful enough and took a big sip. Tea was too hot so that it just *literally* burnt my tongue. Aristotle would probably say that this is a clear example for perception. There is really an object of

⁴⁸ Aristotle, *De Anima*, 419a10-25, 418a29-418b4. Aristotle here discusses the necessity of medium of sensation without which a sense organ will be unable to perceive anything. Air, light, water, etc., are all mediums of sensation. For instance, Aristotle says that nothing is visible without light. For an extensive interpretation, see Aquinas, pp.135-138. Also see Aristotle, *De Anima*, 424b16-18; Aquinas, p.174, where Aquinas comments "563: smelling happens when a thing is affected by odor in such a way as to perceive it. But air is not affected in this way, since it has no sensitive potency; it is affected only so as to become a sense-object, inasmuch as it affords a medium for sensation".

⁴⁹ Aristotle, *De Anima*, 417b23-26. "So it lies in man's power to use his mind whenever he chooses, but it is not in his power to experience sensation; for the presence of the sensible object is essential." Cf. Shields. Aristotle's Psychology. Shields did not add the first clause as a requirement, which I think it is unfortunate to the extent that Aristotle makes a distinction between mind and sensation over that requirement in addition to the distinction between sensible and intelligible form-reception capacities. That is to say, if one does not make use of this requirement in one of the clauses for the formulation of perception, it seems like one of the important points about Aristotle's understanding of sensation and thinking will not be captured by that formulation.

⁵⁰ I take medium of perception as one of the necessary components taking part in building perception of objects. However, Shields does not add this requirement in his formulation of perception. See Shields, Aristotle's Psychology.

perception, in this case tea; and I am a being capable of receiving the sensible qualities, e.g., hotness, coldness, sweetness, color, etc., of tea by means of one of my sense-organs, my tongue in this case, up to the task; and I am interacting with that object of perception through a medium of perception. In this scenario, as a result of the interaction with hot tea, my tongue is made like hot too. But the whole idea embedded in this scenario gives rise to a very challenging suspicion whether the same thing can be said for the other sense organs like eyes, ears, etc. When I perceive a red apple, do my eyes literally become red or contain redness? This seems to be a very difficult point in making sense of Aristotle's idea of perception; and things get much more worse, if he really considers the relationship between agent and patient as a literal resemblance or direct resemblance, in his explanation of thinking.⁵¹

⁵¹ On a related issue, it is also extremely difficult to understand Aristotle's use of "Form" in an actual object and "Form" in mind. See Richard A. Watson, *Representational Ideas: From Plato to Patricia Churchland* (Dordrecht; Boston: Kluwer Academic Publishers, 1995), pp. 7-10. In sum, Watson remarks the following: Aristotle's notion of Form, as being different from that of Plato, tells us that Forms do not have any independent existence. They "have being only in union with matter or a mind" (p. 8). This, however, gives rise to an important issue, says Watson. It seems here we have two different modes for Forms. "In its actual existence in matter, the Form makes the thing what it is" (p. 8). So, given that a thing is composed of a Form "which is the structure, pattern, or plan of a thing" and a particular matter "that has the potential to be formed," then the Form of a thing, for instance, a Hermes statue, is the defining and informing aspect of a thing. But, Aristotle also talks about Forms as objects of thinking in the mind. So, "in its intelligible existence in the mind, that same Form does not inform the mind in the sense of making the mind take on and exhibit in actual existence the plan or pattern of the Form" (p. 8). So, according to Watson, we have two different modes of Form in Aristotle's metaphysics: the informing Form in the actual object and the non-informing Form in the mind. A couple of questions are in order. First, it seems that Forms are really very strange entities because the same form "can have two radically different modes of being at the same time" (p. 9). Second, since the non-informing Form is not like the informing Form regarding the aspect of making something as what it is, then what is its function in the mind? Is this something like a "pure, unexemplified pattern or plan," and "how could we know it?" (p. 8).

Imagination and Thinking

After having explained necessary and sufficient conditions which, when obtained, produce activity of perception in the patient object, Aristotle then moves on to introduce the power of imagination that has an intermediary role between sensation and thinking. Aristotle defines this faculty as that which provides a store of *phantasms*, mental pictures/ images as the effect of movement/ change on a sense organ receiving sensible forms of objects.⁵² One can think of the relation between sensation and imagination as a sense-movement in the former, in which a neutral state of a sense organ moves from its neutral state to an active perceptual state, and an image-movement in the latter, in which sensible forms, images of the objects previously sensed without their material features, fill a potential state of imagination.⁵³

The decisive point about imagination, I shall explain in passing, is that it does have a very crucial task in Aristotle's psychology. On the one hand, imagination always implies previous sensation. That is to say, if one imagines something now, that is only because either he is currently perceiving something and those sensible forms of objects are immediately stored in the power of imagination, or he had a perceptual experience of that object before the time of imagining now. On the other hand, imagination is itself implied by judgment and thinking. Aristotle here claims that as sense objects are necessary for sensation,

⁵² Aristotle, *De Anima*, 428a1-5; 428b10-17.

⁵³ For this interpretation, see Aquinas, pp. 201-203.

phantasms/mental images are necessary for intellect.⁵⁴ This will become clearer after the presentation of Aristotle's account of thinking. So, let us now dwell on his account of thinking.

It is necessary then that mind, since it thinks all things, should be uncontaminated, as Anaxagoras says, in order that it may be in control, that is, that it may know; for the intrusion of anything foreign hinders and obstructs it. Hence the mind, too, can have no characteristic except its capacity to receive. That part of the soul, then, which we call mind (by mind I mean that part by which the soul thinks and forms judgments), has no actual existence until it thinks. So it is unreasonable to suppose that it is mixed with the body; for in that case it would become somehow qualitative, e.g., hot or cold, or would even have some organ, as the sensitive faculty has; but in fact it has none.⁵⁵

Aristotle construes the intellectual power of the soul in two different aspects, one of which he calls "the potential/patient mind" and the other he calls "the actual/agent mind".⁵⁶ The patient mind indicates a certain state of being passive to an intelligible object. Understood in this manner, this is an aspect of mind that is in potency to know everything there is to be known.⁵⁷ How is it possible for the patient mind to know everything? On the issue of potential sensation, Aristotle earlier asserted the claim that the power of sensation must be in a neutral state before any object, through a medium of perception, begins to interact with that power. If

⁵⁴ Aristotle, *De Anima*, 431a14-17; Aquinas, p. 230.

⁵⁵ Aristotle, *De Anima*, 429a21-27. See also 429b0-6.

⁵⁶ In the literature on Aristotle's *De Anima*, there is a very popular convention to understand the mind as of two different kinds: "the agent mind" and "the patient mind". However, I would like to make it clear, at least in a footnote, that I do not intend to use these two terms in order to refer to two different kinds of mind. Aristotle never used these terms; instead he says "the intellect in actuality" and "the intellect in potentiality". Understood in this manner, these two must be understood as two different states of mind, not two different kinds of mind. I owe this point to Chryssi Sidiropoulou.

⁵⁷ Aquinas, p. 205.

my sense organs, say my eyes, already contain blue qualities, then this will prevent them seeing green qualities.⁵⁸ The same sort of reasoning runs here in the case of patient mind as well. The patient mind, like the pupil of eye that is able to receive colors but is itself colorless in potency, is capable of receiving all intelligible forms of objects and it must lack all those things which of its nature it understands and knows.⁵⁹ Had the intellect been restricted to one particular nature, i.e., entertaining only one specific intelligible form, it would have been prevented knowing and understanding other intelligible forms. Aristotle then concludes that the patient intellect is never a bodily thing, nor compounded of bodily things, because those would make its nature of a particular kind and thus prevent it from knowing and understanding other things. This is the reason why the intellect has no bodily organ.⁶⁰ This, again for Aristotle, makes it clear why sensation and intellect differ with respect to their receptive capacities. Upon a violent stimulus, a sense organ may be completely or temporally damaged. For there is a limit for the receptive capacity of each sense organ, and anything above or under that limit is outside its receptive capacity. However, the same is not true for mind. It does not matter whether the intelligible form is very high or low, because there is not a mind organ, one that is designed to understand only within the boundaries of its physical/structural capacities. According to Aristotle, “when mind thinks the highly intelligible”, unlike the case in

⁵⁸ Ibid., p. 135.

⁵⁹ Ibid., p. 206. This is Aquinas’ reconstruction of Aristotle’s argument.

⁶⁰ Ibid., p. 207.

sensation, “it is not less able to think of slighter things, but even more able”.⁶¹

As Aquinas nicely puts it, the patient mind is like “a sheet of paper on which no word is yet written, but many can be written”.⁶² But then what does Aristotle mean by the agent mind? On this point, Aquinas counts three essential characteristics that both the patient and agent mind have in common. These are the features like their separation from matter, their impassibility and their purity (unmixed with the body). Then, what is the essential attribute of the agent mind distinguishing it from the patient mind? The agent mind is essentially in act, and this is its proper attribute.⁶³ This does not make it some kind of separate substance. For Aristotle, the agent and patient mind are only two aspects of the same power of the soul, i.e., mind. The agent mind is essentially in an active state of being that immaterializes and abstracts its objects of thought.⁶⁴ In other words, it abstracts all the particular features of phantasms given by imagination and makes them objects for its own activity of thinking.

⁶¹ Aristotle, *De Anima*, 429b3-6.

⁶² Aquinas, p. 217. This, however, should be understood only metaphorically. In Aquinas’ interpretation, it seems like there is a confusion as if there can be a state of mind in full potentiality that is able to think about everything and be in a completely unaffected state. In this sense, it is really very hard to find an example for the intellect in potentiality. Even when a person is born, we can still talk about a minimum actuality for the intellect. So, there is probably no such thing as the patient intellect *per se*. There seems to be always a minimum actuality. I thank Chryssi Sidiropoulou for helping me understand this point about the relation between potentiality and actuality of the intellect.

⁶³ *Ibid.*, p. 220. Aristotle also compares the agent mind to light. Light makes potential colorful objects visible and actual, and perhaps he wanted to say that, likewise, the agent mind makes the potential intelligible objects thought and actual.

⁶⁴ Aquinas, p. 220. Here, Aquinas explains the reason for postulating the agent mind: “The reason why Aristotle came to postulate an agent intellect was his rejection of Plato’s theory that the essences of sensible things existed apart from matter. For Plato there was clearly no reason to posit an agent intellect. But Aristotle, who regarded the essences of sensible things as existing in matter with only a potential intelligibility, had to invoke some abstractive principle in the mind itself to render these essences actually intelligible.”

Regarding the relation between imagination and intellect/mind, Aristotle puts forward the idea that phantasms/representations/mental images in the power of imagination are necessary for intellect/mind, similar to the case where sense objects are necessary for sensation.⁶⁵ According to Aristotle, sense organs receive sensible forms of objects and leave out their matter in the activity of sensation. Phantasms are, then, something retained in the power of imagination, which are originated from the effect of actual sensation.⁶⁶ Phantasms are the objects of thought and they have still their individuating conditions. For Aristotle, mind works through its power of abstraction, and intellectual notions about phantasms are abstracted from their individuating aspects.

In this context, Aristotle is very much committed to an empiricist idea that one cannot think without experience or mental correlates for that experience. There is always something like a mental picture/image of object in imagination that makes it possible to think the intelligible form of that object. This point makes it difficult to understand how Aristotle would have conceived God. Perhaps, the intellectual notion of God is something that is abstracted and organized around certain sorts of phantasms reminiscent of god-like properties. But, if God were defined as a pure thinking being, then it would be quite troublesome to understand, without an experience of such a being, how someone can be able to think of God?⁶⁷

⁶⁵ Aristotle, *De Anima*, 432a3-10.

⁶⁶ Aquinas, p. 235.

⁶⁷ Aristotle constantly emphasizes the thesis that mind cannot operate its intellectual power without mental correlates/phantasms that emerge as the effect of perceptual activity. However, if the activity of thinking is always processed over the phantasms stored in imagination and imagination always relies on sensation, then what would be

Aristotle defends the view that soul is the form of body as an essential component of an individual thing. One might be tempted to conclude, after having witnessed the tension between mind as a separable power of the soul and other powers of the soul that definitely involve body, that mind reveals completely different characteristics of a soul and this must be an evidence that mind is a distinct substance. Following Aquinas' interpretation, I think that that sort of conviction is fundamentally against what Aristotle would have wanted to invoke from the very beginning of *De Anima*. Soul is the form of body and it is always embodied. Patient and agent minds are only distinct powers of the soul and they are never to be understood as distinct substances on their own. When Aristotle says that they are both separable from the body, he only wanted to show that mind must be free from the bodily organs. Otherwise, it would have a nature of a particular kind, something that has receptive limitations about that particular kind. In that case, it would make impossible to explain, what most of us believe, the universal capacity of mind.

But, although the agent mind and the patient mind do not refer to a distinct substance, it seems that Aristotle really believes that the agent mind is immortal and perpetual.⁶⁸ So, there is a sense in which, besides being free from the bodily organs, the agent mind shows a radically different sort of characteristics from that of the patient mind. The patient

like Aristotelian thought of God? In other words, since we do not have a sensual experience of God as something existing in our environment, then it is clear that we will not be able to think of God. If it is true, like Aristotle says, that the soul never understands apart from phantasms and never thinks without a mental image/representation, then, given that it is possible to think of God, I wonder what would it be like to have an Aristotelian thought of a god?

⁶⁸ Aristotle, *De Anima*, 430a18-26.

mind is always related to the body, in a sense, it signifies the passive aspect of having a mind during the experience of this world. That's why Aristotle believes that the patient mind is perishable.⁶⁹ When a human body dies, so dies its patient mind. However, the agent mind is a part of the soul whose activities are proper only to itself, and it is, for this reason, completely independent of body. Therefore, Aristotle says, the agent mind is immortal and perpetual.⁷⁰

As for thinking (and knowing), Aristotle almost carries on the same attitude, the explanatory attitude exhibited in his account of sensation, in the case of thinking. There is an object of thought (the object does not need to be present in my environment) acting upon my capacity receptive of certain sorts of intellectual forms by means of *phantasms* in the power of imagination. The mind simply abstracts individuating features of *phantasms* and provides universal intellectual notions for objects of thought. As a result of this intellectual communication between imagination and mind, my mind becomes like that object of thought outside in the environment with respect to its intelligible form. Here is the formulation of Aristotle's account of thinking:

A subject, S, thinks an object, X, iff (1) S is able to receive the intelligible forms of X,⁷¹ (2) X acts upon sensation and sensible qualities/*phantasms* of X are stored in the power of imagination, (3) particular properties of those *phantasms* are abstracted by the power of mind and (4) that receptive capacity of intelligible forms, i.e., mind, resembles X with

⁶⁹ Aquinas, p. 222.

⁷⁰ Aquinas, p. 222. This is Aquinas' interpretation. Aristotle does not really give an argument for the survival of the agent mind over death of a body.

⁷¹ As far as I conceive, Aristotle thinks that the presence of the object of perception is essential for the perception; but he does not make the same claim in the case of thought which means that one can think in the absence of object of thought in the perceptual field. Compare my formulation here with the formulation in Shields, Aristotle's Psychology.

respect to its intelligible forms (S's intellectual capacity becomes isomorphic with the intelligible forms of X).

Let's try to comprehend, in a following simple example, what is involved in this Aristotelian cookbook about thinking. This morning, I had the painful experience of drinking a cup of hot tea. I began to think, after this experience, that tea was hot. In accordance with Aristotle's conception of thinking and mind, then I was in an intellectual interaction with the intellectual forms of those objects, e.g., tea-form and hotness, of my thought and my mind gradually resembled the intellectual forms of those objects of thought, e.g., being tea and hot. Of course, my example is the one based on an experience, but the situation does not get any better if I focus on thinking about an object of thought that is not supported by any experience. While I think of the statement that God is omniscience, how is it possible for me to share the same form with being God and being omniscience? It is quite difficult to understand, if Aristotle really means the relation of literal resemblance, how my mind can resemble the forms of my object of thought in the activity of thinking. The same kind of difficulty also applies to perception. That is to say, does Aristotle really believe that resemblance literally takes place between the object and perception of that object? Some commentators, for instance, took Aristotle to imply something very simple and literal/real resemblance.⁷² On this approach, when I see something red, then my eye really becomes red. "So, *likeness* amounts to shared-property exemplification".⁷³ However,

⁷² Richard Sorabji, "Body and Soul in Aristotle," *Philosophy* 49, no. 187 (January, 1974), pp. 63–89. He interprets Aristotle as favoring this literal/real resemblance.

⁷³ Christopher Shields, "Controversies surrounding Aristotle's Theory of Perception." In *Aristotle's Psychology*. Italics belong to Shields.

some commentators pointed to passages where Aristotle seems to be committed to non-literal understanding of resemblance. On this approach, my sense organs or mind resemble their objects without exemplifying the sensible forms perceived or intellectual forms thought.⁷⁴ Consider the following example about the relation of resemblance between a house and its blue print. So, for instance, when an architect points to the blue print and says “this one is a house”, he does not actually mean that the real house and its blue print exemplify the property of being a house; but he means that the blue print “somehow *encodes* that property”.⁷⁵ Following this line of thinking, some commentators believe that Aristotle has in mind a non-literal resemblance when he talks about perception and thinking. On this alternative interpretation, when I see a red object or think of redness, my sense organs or mind do not exemplify the sensible or intellectual form of being red. They do not really become red, but they somehow encode those forms like in the case of a blue print of the house encoding the structure of being a house.⁷⁶

In the end of this chapter, I would like to present Aristotle’s explanation for the occurrences of misrepresentation. Firstly, as far as I understand Aristotle, in order to mentally represent something there must be both a causal and (some sort of) resemblance relationship between the external object and mind. Since imagination and intellect both depend on experience, then Aristotle would probably say that we

⁷⁴ Ibid.

⁷⁵ Ibid. Italics and the example belong to Shields.

⁷⁶ Ibid.

wouldn't be able to represent anything in the absence of the causal interaction with the world. So, without causality there would be no mental representation of the world. The same also goes for the relation of resemblance, namely, without a resemblance relation between the world and mind there would be no mental representation of the world. Only when these two general conditions, among other specific conditions, are satisfied, we will be able to represent the world. This is, I take it to be, the Aristotelian reason why we do have mental representations about the world. Nevertheless, it is quite possible that our mental representations occasionally give us a wrong picture about the world, how things actually stand in the world. So then, I wonder how would Aristotle explain this aspect of representing capacity? For instance, what would Aristotle say for the case when I mentally represent a dog when the cause of that mental representation is actually a wolf in the dark? There is a cause of mental representation (e.g., wolf) and there is a degree of resemblance between the represented object/cause (wolf) and representation (dog). I believe we can find something in *De Anima* where Aristotle, at least, gives us some crucial insights about his possible response. As far as I see, there can be two ways of explaining the occurrence of misrepresentation in Aristotle's account; one comes from the power of imagination and the other comes from the non-ideal external/internal conditions for the power of sensation. Let's begin with imagination. In Book 3, division VI, Aristotle begins explaining what imagination is in itself and its properties:

But since it can happen that, one thing moving, another is moved by it; and *imagination seems to be a movement*, and to arise only with sensation, and in sentient beings, and to be of

such objects as are sensed; and since a motion may be caused by actual sensation, and such *necessarily resembles sensation*,—then *imagination* will be just this movement, never originated apart from sensation, incapable of existing in non-sentient beings, and enabling its possessor to act and to be affected in many ways, and being itself both *true and false*. This happens because sense-perception is *true* of its own *proper objects*, or has the least amount of *falsehood*; but secondarily it bears on that in which these qualities inhere, and here it can be deceived. Sensation is reliable as to whether a thing is white or not, but not as to whether it is this or that. Thirdly, there are the *common sensibles* consequent on the accidents in which the proper qualities inhere. I mean, such as movement and dimension, which belong to sense-objects—and about these deception very easily arises in sensing. The movement derived from actual sensation differs from the sensations by which these three objects are perceived. Although the first [movement] is true, the sensation itself being present, the others can easily be false, whether sensation be present or not, and especially when the *sense-object is distant*... [I]magination is a movement produced by sensation actuated... And since these images *dwell within*, and resemble sense experiences, animals do many things in accordance with them; some animals, as lacking reason, namely beasts; but some, i.e., men, when their intellect is veiled either by *passion*, or by *sickness*, or by *sleep*.⁷⁷ (my emphasis)

In interpreting this long passage, Aquinas says that “imagination is a sort of movement” and just as we can analyze sensation as the psychophysical⁷⁸ movement of a subject by sensible things, we can also think of the activity of imagining as the psychophysical movement of a subject by “certain appearances called *phantasms*”.⁷⁹ But why does this movement in imagination resemble the movement in sensation? This has

⁷⁷ Aristotle. *De Anima*, 428b10-429a9. Here I use the translation of the text in St. Thomas Aquinas. *Commentary on Aristotle's De Anima*, translated by Kenelm Foster and Silvester Humphries (Notre Dame, Indiana: Dumb Ox Books, c1994), p. 201.

⁷⁸ There is a reason for using the term “psychophysical” here. For a nice discussion of this point within the context of the controversies whether Aristotle was a functionalist or not, see Chryssi Sidiropoulou (forthcoming), “A New Understanding of Aristotle's ‘Dualism’ in the *De Anima*,” *Philosophical Inquiry*.

⁷⁹ Aquinas, p. 201. My emphasis.

to do with Aristotle's commitment to the general principle that "a thing moved may move another" and "that which, being moved, moves another, must cause a motion similar to its own".⁸⁰ So, the idea here seems to be the following: when we sense something, we are in a psychophysical movement whose cause is some sensible object. Through this process, the relevant sense organ acquires the sensible form of objects by which we are moved. This perceptual process/movement then leads to image-movement in imagination, and we have something similar (phantasm/appearance/representation) to the sensible form of the object. Furthermore, imagination "cannot exist without sensation".⁸¹ This sums up the first half of the passage. Another question, however, is in order here: if imagination always depends on sensation, and sensation has the least amount of falsehood, then why does Aristotle believe that imagination, slightly different from sensation, is sometimes true and sometimes false? Does not it follow that if imagination resembles sensation, then we should talk about few cases of falsehood for both sensation and imagination? Not quite, Aristotle would have said. He would probably say that something is lost on the way from sensation to imagination, and we should observe the details. First of all, we should keep in mind that the source for the imagination comes from "the act of exterior sense" and this "is related to truth in different ways, according to its *varying relation* to objects".⁸² Then, what is its varying relation to objects? First of all, about proper sensibles (e.g., color) sensation rarely

⁸⁰ Ibid., p. 202.

⁸¹ Ibid., p. 202.

⁸² Ibid., p. 202. My emphasis.

deceives us, Aristotle believes. It is the natural power of our eyes to report, among other proper objects of sensation, colors of the objects, and we rarely fail in sensing the colors, whether this white thing is really white or black. “Thus only in a minority of cases do senses judge inaccurately of their proper objects.”⁸³ For instance, when I am sick I may not be able to taste the things as I do under normal conditions, because my tongue is being affected by sickness. So about proper objects of my sensation, I can be deceived only rarely when some internal defect is the issue with one of my sense organs. This is the first case, internal non-ideal condition, for my sensation to go wrong. As for the second case, Aristotle mentions indirect objects of sensation. We do not err in our sensation of white things, “what seems to be white”, under normal conditions, “is indeed white as the sense reports;” but when it comes to judging whether this white thing is snow or floor, our senses often deceive us, “especially at a *distance*”.⁸⁴ So, this is the second case, external non-ideal condition and indirectness of sense object, for my sensation to go wrong. As for the third case, Aristotle talks about the deceptive character of common sensibles (e.g., size, movement, magnitude, etc.) “found in things some of whose accidental qualities are proper sensibles”.⁸⁵ The case seems to be the following one. When we sense something, first we sense its proper sensibles, whether it is white or black, whether it smells like something or other. In these cases, we rarely fail in sensing the proper sensible forms of objects. Also, “size and movement accompany, as common sensibles,” the

⁸³ Ibid., p. 202.

⁸⁴ Ibid., p. 202. My emphasis.

⁸⁵ Ibid., p. 202.

proper sensible forms of things.⁸⁶ For these kinds of cases, “our judgment has to adjust itself to differences of distance, things seen further away seem the smaller”.⁸⁷ So, for instance, when I look at a ship from the coast, my senses can deceive me in judging that it is a small object. With this last case, Aristotle completes his list on the ways sensation can go wrong.

Then, why is there more falsehood in imagination than in sensation, given that they resemble each other? I shall quote Aquinas here:

Now the movement of imagination, being derived from the actuated senses, differs from these three types of actual sensation as an effect from its cause. Thus, just because effects, as such, are weaker than their causes, and the power and impress of an agent is less and less evident the further away are its effects, therefore imagination is even more liable than are the senses to fall into error which arises from a dissimilarity between the sense and its object... It follows too, from what we have said, that imagination is generally truthful when it arises from the action of the ‘proper sensibles’; I mean, at least, so long as the sensible object is present and the image-movement is simultaneous with the sense-movement.⁸⁸

This is very close to the British empiricist idea that ideas are less lively copies of sense impressions. Consider the following case: The object of sensation is still present, what I am sensing is its proper sensible form, and I am imagining it through its effect as an appearance/representation/phiasm in the power of imagination. In these kinds of cases, imagination is generally truthful. But, when the object is not present and I started to lose contact with it, the connection

⁸⁶ Ibid., p. 202.

⁸⁷ Ibid., p. 202.

⁸⁸ Aquinas, pp. 202-203.

between sensation and imagination gets weaker, and hence I start to imagine things very loosely and my imagination can deceive me, “even with regard to proper sensibles”.⁸⁹ “We may imagine *absent* black things as white”.⁹⁰ This is the case for my imagination to go wrong, even concerning proper sensibles. The situation gets worse when “the image-movement arising from perception of objects indirectly sensed, or of the common sensibles”.⁹¹ Here, it does not matter whether the object is still present or not. Imagination, in these kinds of cases, often goes wrong. When the object is absent, it is “more liable to error”.⁹² Last but not least, Aristotle emphasizes that “images dwell within” even when the object is no longer present, so there must be some significance of having those images in the imagination as to determine the behavior of animals. Aquinas says “just as sensations arouse appetitive impulses whilst the sensed objects are present, so do the images when things are absent”.⁹³ Some animals, due to the lack of intellect, can imagine things and, in order to fulfill their desires of imagination, may do crazy things. The mind in men, however, enables people to command imagination and keep it under control. Of course, Aristotle says, there are certain kinds of circumstances that imagination takes over control and then a lot of falsity would come out of imagination. For instance, when somebody is in love or angry, his/her reason may go to vacation and the person may start to lose his/her grip with the reality that is being represented, under normal

⁸⁹ Ibid., p. 203.

⁹⁰ Ibid., p. 203. My emphasis.

⁹¹ Ibid., p. 203.

⁹² Ibid., p. 203.

⁹³ Ibid., p. 203.

conditions, with the aid of governing reason. Let this suffice for the cases where imagination misrepresents things about the world. In summary, I think Aristotle (roughly) would say, for the cases of misrepresentation, that there can be two ways where we begin to misrepresent the world. First, when the conditions, both internal and external to sensation, are not ideal, our sensation can deceive us. So, in the case of my mental representation of wolf (in the dark) as a dog, Aristotle would have said, "it is not ideal for sensation to work properly here". Second, when the connection between sensation, imagination and intellect (among other factors) gets weaker in the process of mental representation, then imagination can deceive us. I may immediately start associating, in my imagination, the wolf-image with something similar to it, for instance, the dog-image. My intellect can be veiled by my passion of dogs, so I may represent a wolf as a dog. I may start representing everything around me, for instance when I am mentally deluded or dreaming, as dogs; I may represent the world in which human beings are replaced by puppies.

Let's recapitulate the main ideas in this chapter. Aristotle chiefly maintains that intrinsic capacity of form-reception, causal relationship between the patient and agent, and resemblance, together with other specific conditions, provide the ground for perceptual and intellectual states of the soul. Without these conditions satisfied, we can have no representation of the world, and misrepresentation of the world may occur either due to the non-ideal internal/external perceptual conditions or, among other things, the weakening of the connection between sensation, imagination and mind.

CHAPTER III

INTRODUCTION TO METAPHYSICS OF MIND

The main target of this chapter is to introduce and critically assess three different theories of mind in contemporary philosophy of mind that have been offered to explain the nature of mental states. For this purpose, in the first section, I will begin with the short sketch of behaviorism and then comes the brief overview of the mind-brain identity theory of mind in the following section. Afterwards, in the third section, I will get into the technical details of functionalism, the theory of mind that has been the cornerstone of contemporary philosophy of mind since its first debut. In particular, I will highlight its theoretical virtues and great advantages over behaviorism and the mind-brain identity theory. In the last section, I will attempt to reveal the fact that there is a hidden Platonic ideology behind functionalism, and I will claim that there can be a rather different way of reading functionalism that may prove to be much more sensible and fruitful.

Behaviorism

Having its historical roots back to J. B. Watson's revolutionary perspective and works in the beginning of the twentieth century, behaviorism made an appearance in the psychology literature as a methodological reaction to "the subjective and unscientific character of

introspectionist psychology”,⁹⁴ and, alternatively, as a new way of conducting psychological experiments. According to Watson’s conjecture, it was impossible to derive any scientific information from a set of heavily unreliable and extremely subjective data about mental life and consciousness. Therefore, psychology, as a branch of science, must give up dealing with the psychoanalysis of inner mental episodes and consciousness, and it must be concerned with only “publicly observable human and animal behavior”.⁹⁵ Influenced by Pavlov’s stimulus-response model of explanation, Watson suggested that instead of taking conscious experience as the proper object of psychology, we must develop, in opposition to the introspectionist approach, a new research program called “Behaviorism”, which should investigate the observable facts about the adjustment of organisms, man and animal alike, to their environment and the connection between stimuli and responses of those organisms in a given environmental condition.⁹⁶ In this respect, the proper objects of psychology, Watson and other behaviorists firmly believed, must be something that is objective, testable and measurable. As thus understood, the initial goal of psychology is to discover general laws of behavior by means of which one can control behavior and make

⁹⁴ Jaegwon Kim, *Philosophy of Mind* (Boulder, Col.: Westview Press, 1996), p. 25. In this section, my primary goal is to present a very concise overview of philosophical behaviorism. Therefore, I will leave out the discussion of methodological behaviorism, psychological (radical) behaviorism and behaviorism as a socio-political theory. For methodological behaviorism, see John B. Watson, "Psychology as the Behaviorist Views It," *Psychological Review* 20, no. 2 (1913), pp. 158-77. For psychological behaviorism, see Burrhus Frederic Skinner, *Science and Human Behavior* (New York: Free Press, 1965). For behaviorism as a socio-political theory, see Burrhus F. Skinner, *Walden Two* (New York: Macmillan Co. 1962). For the real time influence of Skinner’s behaviorism as a socio-political theory in *Walden Two*, see various online and printed materials about Twin Oaks Community in Virginia in the U.S.A.

⁹⁵ Kim, p. 25.

⁹⁶ Watson, p. 167.

exact predictions about behavior.⁹⁷ Behaviorism, later to flourish and be called “behavioral science”, was found to be successful in explaining a wide range of human and animal psychology, especially about memory and learning history.⁹⁸

On the other hand, behaviorism, as a philosophical doctrine, shares something in common with its closely related cousin “psychological behaviorism”. Both behaviorist movements objected to the Cartesian metaphysics, in which the mental is understood as something “essentially private and subjective”,⁹⁹ on the grounds that knowing other minds turns into a big astonishing enigma. In this theory, only a single Cartesian soul/mind has direct and privileged access to the perceptual and cognitive occurrences in his own mind, and, for this reason, the rest of society of souls/minds has no conceptual or empirical evidence/tool to rely on, except his verbal and non-verbal behavior, so as to ascribe experience and thinking to him. The problem with this view, many philosophers and psychologists maintain, is that the Cartesian notion of mentality simply makes it impossible to know whether a being/system is a brute automaton or it has a mind.¹⁰⁰ Much like in Wittgenstein’s example of “the beetle in the box”, each soul/mind is only permitted to know/look into his own mind, and does not have the remotest idea of what other souls are thinking or perceiving, and what is worse, he could not know whether there exist other souls. This gives rise

⁹⁷ Ibid., p. 158.

⁹⁸ Kim, p. 25, pp. 41-44.

⁹⁹ Ibid., p. 26.

¹⁰⁰ Ibid., p. 26; Keith Campbell, *Body and Mind* (Notre Dame, Ind: University of Notre Dame Press, 1984), p. 63.

to a very difficult question: How could anybody ever come to understand the meaning of the word “pain” and be able to express his pain in a society of mind successfully? The Cartesian metaphysics of mind transforms this very basic and obvious fact about social communication into a big riddle. This was unacceptable, for a great number of people. Behaviorists were probably tired of the Cartesian world where the magic tricks have been used endlessly / abundantly and they could not bear with more magical ideas. The idea of a single Cartesian soul, divided into two, the former of which is engaged in sensing and thinking about something, the latter of which enjoys the privilege of being the only audience to watch the play of sensations and thoughts performed by the former, was a big negative stimulation for the behaviorist community.¹⁰¹

In opposition to the Cartesian idea of an inner private theatre, behaviorism puts forward the claim that each mental phenomenon, however the degree of complexity, can be explained “by reference to facts about publicly observable behavior”.¹⁰² So, for behaviorists, any mental state, intentional or phenomenal, is a disposition, or better, family / complex of dispositions, to exhibit a certain kind of behavioral patterns in response to a certain kind of stimulus conditions.¹⁰³

Accordingly then, to have a fear (of some sort) is just to acquire a

¹⁰¹ Kim, pp. 26-27. The idea is commonly known as “Cartesian inner private theatre”.

¹⁰² Ibid., p. 28.

¹⁰³ Gilbert Ryle, *The Concept of Mind* (London: Penguin, 2000 Originally published: London: Hutchinson, 1949), p. 43. Ryle states that possession of a dispositional property “is not to be in a particular state, or to undergo a particular change; it is to be bound or liable to be in a particular state, or to undergo a particular change, when a particular condition is realized.” Cf. Armstrong, David. “The Nature of Mind.” In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980a), p. 194.

complex disposition; and acquiring a disposition, in one sense of the term “disposition”, “is just having come true some conditional statements describing my tendency to behave”.¹⁰⁴ The following example is a case in point:

A subject, S, is in fear of flying (in the state of aerophobia)=_{def.} If S is seated in an airplane ready to take off, then S, under normal conditions, will typically exhibit certain kinds of behavioral patterns (e.g. screaming desperately, crying out loud, performing gestures of such-and-such kinds, when asked “What is the matter (with you)?” he will typically respond, “Flying is dangerous and not natural for human beings”, “I do not like flying”, etc.)¹⁰⁵

The example is meant to capture the project of logical / analytical behaviorism. According to this theory, “any meaningful psychological statement, that is, a statement describing a mental phenomenon, can be translated, without loss of content, into a statement solely about behavioral and physical phenomena”.¹⁰⁶ Logical / analytical behaviorism, having its theoretical origins from logical positivism, holds on to the big project of translation through which one will get clear about the implications of the mental-causal talk and explain the mental properties in terms of behavioral dispositions, while, at the same time, avoiding any reference to inner private episodes of individuals. At the end of this translation project, logical / analytical behaviorism hopes to demonstrate, at least in principle, that behavior is not only a sign of mentality, but also it constitutes mentality. So, “having a mind just is a matter of exhibiting,

¹⁰⁴ Campbell, p. 72.

¹⁰⁵ I owe this example to Stephen Voss.

¹⁰⁶ Kim, p. 29.

or having the propensity to exhibit, certain appropriate patterns of observable behavior".¹⁰⁷

What is disposition to behave, and how does it differ from cause of behavior? In passing, I shall say a few words about this difference. Let us go over some examples about disposition first. Consider dispositional properties such as "being soluble in water", "being magnetic" and "being fragile". Take the definition of fragility for instance. Any object O is fragile, iff O is easily broken or damaged, under normal conditions, whenever some suitable external object strikes/hits O. In a similar fashion, behaviorism defends a dispositional account of mental states. For instance, being angry, for a behaviorist, does not signify an inner mental episode that causes angry-behavior patterns; rather, it is a disposition¹⁰⁸ or complex of dispositions to exhibit angry-behavior patterns when the circumstances are normal and right set of stimuli obtain. But, why does a behaviorist believe that a mental state, anger say, is not the cause of angry behavior? First, behaviorism commits to the idea that "the connection between mind to behavior is too close to be causal".¹⁰⁹ For instance, writing a poem, as a behavior, does not really indicate the ending of mental activity for which the behavior is the end-result.

¹⁰⁷ Ibid., p. 26. This statement might be a source of confusion to the extent that it gives rise to the suspicion that all behaviorists say that any given mental state is either behavior or behavioral disposition. However, the statement only stands for the different approaches taken by early behaviorists and later behaviorists. Early behaviorists claim that the mental states are essentially publicly observable behaviors associated with those mental states. Later behaviorists like Ryle, on the other hand, identify mental states with dispositions to behave. So, later behaviorists can give an account of cases where a mental state is not exhibited in a behavior.

¹⁰⁸ Campbell, p. 65.

¹⁰⁹ Ibid., p. 65.

“Writing a poem ... is itself a piece of mental activity”.¹¹⁰ Campbell nicely illustrates this aspect of behaviorism:

What is out of the question is that mental events, processes, or conditions should play a causal role in producing the behavior which is a manifestation of *that* mental event, process, or condition. To call the behavior a *manifestation* of the mental state is already misleading. The behavior *is* the mental state,¹¹¹ to the extent that anything categorical constitutes a mental state. The mental state is never a cause of its own behavioral elements, just as nothing is cause of itself.¹¹²

For behaviorism, it is a fundamental error to lead into the direction of common thought and ordinary language in which mental events are always taken as the causes of behaviors. The fundamental error, for behaviorism, seems to be grounded on the wrong attitude to analyze mental descriptions. Take the sentence “He built a house”. From this sentence, one is easily led to believe that there is “a sequence of public events in a public space involving physical rearrangements”.¹¹³ Now, take the sentence “He built a fantasy”. This time, people wrongly think that the statement describes “a sequence of private events in a private space

¹¹⁰ Ibid., p. 65. I borrowed Campbell’s example about writing a poem.

¹¹¹ It does appear to me, contrary to Campbell’s interpretation, that it is wiser to identify mental states with dispositions to behave rather than with actual behaviors.

¹¹² Ibid., p. 66. Italics belong to Campbell. The point is related to the distinction drawn between reasons and causes. Before Davidson’s seminal essay “Actions, Reasons and Causes”, the Wittgensteinian approach was very orthodox. According to this approach, explaining action with reference to reasons indicate different kinds of descriptions under which action can be evaluated as “intentional” or “non-intentional”. This way of explaining the phenomena of intentional actions radically differs, according to Anscombe, from the way of explaining the causation between two events, because while the former depends on conceptual/logical relations, the latter depends on contingent/causal relations that hold between two events. As opposed to this tradition, Davidson developed the argument in order to demonstrate that rational explanation is a form of causal explanation. In this regard, his essay was a turning point in the twentieth century philosophy of mind that helped change the way mental states were understood back then. See Donald Davidson, “Actions, Reasons and Causes.” In his *Essays on Actions and Events* (Oxford: Clarendon Press; New York: Oxford University Press, 1980). See also Gertrude Elizabeth Margaret Anscombe, *Intention* (Cambridge, Mass.: Harvard University Press, 2000).

¹¹³ Ibid., p. 68.

involving mental rearrangements".¹¹⁴ Behaviorists believed that this way of understanding mental events eventually lead one to postulate a spiritual mind for which mental objects like anger, pain, etc., fills its states or serve as its spiritual/ psychological contents. So, instead of leading into a dualist metaphysics, they preferred to understand mental descriptions not as referring to private episodes but to refer to tendencies, or dispositions to behave in a public space.¹¹⁵

So far, I presented a very brief sketch of behaviorism. Now, I would like to discuss some of the famous charges pressed against behaviorism. First of all, it may be true that behavior is the most evident sign of mentality, especially in the case of mental states such as being in pain, being angry, being thirsty, etc. But, it may be completely another matter to describe higher cognitive states. For instance, how can a behaviorist provide a behavioral description of beliefs like "there is no largest prime number"; "an independent judiciary system is essential to a democratic government"?¹¹⁶ These are states that are difficult to describe in behavioral terms. For the explanation of these kinds of states, behaviorists, in general, apply for "verbal behavior". So, a behaviorist description for a belief includes more or less like the following:
*"S believes that p " =_{def.} If S is asked, "Is it the case that p ?" S will answer, "Yes, it is the case that p "*¹¹⁷

¹¹⁴ Ibid., p. 68. Both examples belong to Campbell.

¹¹⁵ Ibid., p. 68.

¹¹⁶ Kim, p. 32. Both examples belong to Kim.

¹¹⁷ Ibid., p. 32.

The difficulty, this time, for a behaviorist is to give a pure behavioral/physical description of verbal behavior without presupposing any psychological notion such as “understanding”, “meaning”, “intending to mean that p”, etc. It does seem, for many philosophers, however, that this is almost an impossible task, because the idea of verbal behavior is actually a cluster term for all these psychological conditions and events. Therefore, behaviorism, for some people, fails to explain at least higher cognitive states.¹¹⁸ If higher cognitive states are part of mentality, and if behaviorism can’t find a way to explain those states, then it follows that the thesis of behaviorism that behavior is constitutive of mentality is wrong. The argument, however, rests on the possibility of giving an account of language without presupposing psychological notions. Nevertheless, this task of giving an account of language without the baggage of psychological notions may not be a completely hopeless and impossible task. Wittgenstein and Ryle, for instance, have a way of explaining verbal behavior that excludes private and introspective episodes one is supposed to experience during thought.¹¹⁹

The difficulty of explaining verbal behavior may seem to be only the tip of the iceberg, and the real challenge might be something larger and insurmountable. As a second, and perhaps the most powerful and

¹¹⁸ Contrary to Kim’s conviction, this argument may not be the reason for drawing people away from behaviorism to functionalism. For both behaviorism and functionalism rely on verbal and non-verbal behavior as a criterion of mentality. One can point out that the best way to understand whether a subject believes that there is no largest prime number is to ask him and wait for his verbal response. The Turing Test, which is generally cited as a functionalist test, can be understood as a behavioral test for mentality as well. I owe this point to Stephen Voss.

¹¹⁹ Ludwig Wittgenstein, *Philosophical Investigations*, translated by G. E. M. Anscombe (Oxford: Blackwell, 1991, c1953); Ryle, *The Concept of Mind*.

celebrated, criticism of behaviorism, comes Putnam's attack on behaviorism.¹²⁰ Putnam invited people to re-consider the behaviorist construal of the relation of behavior and mentality. In Putnam's counter-examples for behaviorism, we have a community of super-Spartans who train themselves, for some ideological reasons, not to exhibit any sort of pain-behavior even though they feel pain. Also, Putnam suggests that there could be perfect actors who do not feel pain when actually hit in the face, either due to some neural condition or due to pain-killing drugs, and yet still can manifest the same patterns of pain-behavior (as we do when we are in pain). At the end, Putnam and many others came to the conclusion that "not only can there be pain-behavior without pain, there can be pain without pain-behavior or any disposition thereto".¹²¹ The argument is taken, for many people, to refute behaviorism.

I tend to believe, nonetheless, that a good behaviorist can accommodate both of these cases within the conceptual borders of behaviorism. Let us recall behaviorism. A behaviorist begins with the definition of "disposition". Any object, O, is fragile, iff, O is easily broken, under normal conditions, whenever some suitable external object hits O. The definition is meant to specify that not all kinds of external objects are eligible to break O. For instance, a feather would not have any effect on O since it does not have enough power to break anything. So, a feather is not a suitable external object to break O. Also, not in all conditions, O can be broken. For instance, the external object, although it may carry enough force to break O, say a piece of glass, may not be able to break the same

¹²⁰ Putnam, Hilary. "Brains and Behavior." In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980a), pp. 24-36.

¹²¹ Campbell, pp. 73-74.

object under different conditions, like in the water for instance. So, we need to specify the conditions (whether normal or not) and the agent object (whether it is eligible object to break anything, and so forth) if we want to produce a definition about fragility. Then, by the same token, a behaviorist can provide a dispositional account of mental states with a special focus on normal circumstances. Any individual / subject, S, is thirsty, iff S has the tendency / disposition to exhibit behavioral patterns (e.g. water-searching behavior, putting his tongue out of mouth, etc.) and when water is available, under normal circumstances, he will be drinking some. Then, there is a good chance that we can observe that S will be exhibiting water-related behavioral patterns under normal conditions (e.g., when there are not any environmental / social pressures preventing the individual from exhibiting his water-related behavior, or when there is not any physical abnormality that makes his having the disposition to exhibit water-related behavior temporarily or permanently unavailable). In the case of super-Spartans, we can observe social pressures imposed on any single individual living in the community of super-Spartans. It is their moral code not to exhibit any pain-related behavior, but Putnam insists that they are still in pain (internally). So, the concept of pain still applies to them. At this point, a good behaviorist can say that it is still possible for any member of this society to exhibit pain-related behavior when, for instance, social pressures are removed from his environment. So, it does not mean that super-Spartans have found a way to shut off their pain sensations completely; they are still in pain and this pain can

be exhibited, for instance, when no super-Spartans are around (that is to say, when the social conditions are normal).¹²²

As for the case of perfect actors, we have this time abnormal physical conditions, conditions like the lacking of pain-fibers or taking pain-killing drugs. So, it may be the case that a perfect actor, due to some neural condition or pain-killing drugs, may undergo a typical pain-causing stimulus during the play and he may produce typical pain behavior, indistinguishable from those individuals who exhibit the same kind of pain behavior. But, if I was a behaviorist, I would say that since the conditions are not physically normal and the same, then even though the actor and others are stimulus-behavior equivalent, the actor is not exhibiting a pain behavior. He only pretends to be in pain, and pretending is only another kind of behavior. The question, of course, for a good behaviorist is to distinguish the act of pretending to be in pain from the real behavioral situation of being in pain. However, this should not be a big problem for behaviorists. If a behaviorist is allowed to observe the behavior of those perfect actors for a longer period, it will be a matter of time for them to detect that there is something physically wrong with those perfect actors.¹²³

For these reasons, I do not really understand how Putnam's curious cases of super-Spartans and perfect actors can refute behaviorism. If they do something, I guess, they could only show us that, if you find abnormal or extreme cases, then you may easily spot an error

¹²² Assuming, of course, that the code is binding for each member of the super-Spartan community when and only when super-Spartans are around.

¹²³ For a full discussion of dispositions, see Alexander Bird, *Nature's Metaphysics: laws and properties*. Oxford: Clarendon Press; New York: Oxford University Press, 2007.

about any general theory. However, if a behaviorist is allowed to incorporate the clause about "normal physical/ social conditions", then I do not see any threat posed by Putnam's counter examples of super-Spartans and perfect actors.

There is really some grain of truth to the behaviorist way of describing mental states. Epistemically speaking, behavioral data is probably the most powerful evidence to ascribe mentality to beings/systems. At one point, when we want to understand the level and the degree of intelligence and consciousness of creatures we interact with, behavior seems to be the first, and arguably the most crucial, touchstone by which judgment about the intelligence and mentality of other creatures is made. Moreover, to distinguish a mental state from another, say "fear" from "excitement", we often apply to bodily actions associated with those mental states.¹²⁴ So, behaviorism, in a way, was in the right track regarding the analysis of correlations that often hold between mental states and behaviors; yet, later it was found to be inadequate to explain the real nature of having a mind, something that is the internal cause of intelligent and conscious behavior.

The Mind-Brain Identity Theory

One of the difficult sides of behaviorism is that it simply counts internal processes and episodes an individual has as irrelevant to explain mentality, and, in particular, it denies the view of mind as something that is the internal cause of behavior. However, with the development of

¹²⁴ Campbell, p. 75. The example about fear and excitement belongs to Campbell.

neuroscience and empirical findings in brain research, people began to think that denial of the brain states and processes as the internal cause of behavior, and thereof mentality is, to put it mildly, very unscientific as well as very counter-intuitive. The mind-brain identity theory, in this respect, was a new and promising theory of mind in the late 1950s that not only replaced behaviorism, but it was also built upon a firm scientific ground.¹²⁵

The mind-brain identity theorists maintain the claim that the mind is nothing over and above the brain, so it is identical with the brain. More explicitly, “to have a mentality is to have a functioning brain of appropriate structure and complexity”.¹²⁶ As the scientific evidence or the premise for this identity claim, identity theorists remark upon the straightforward systematic correlations that hold between mental events and brain processes.¹²⁷ For instance, any kind of brain damage, minor or severe, ends up with partial or complete loss of mental capacities and character traits of a person. Taking drugs and consuming alcoholic beverages are definitely very influential factors determining the emotional and intellectual responses a person may find appropriate in a

¹²⁵ Jaegwon Kim, *Philosophy of mind*, Dimensions of philosophy series (Boulder, Col.: Westview Press, 1996), p. 52. In this section, my target is not to present the historical details about the rise of the mind-brain identity theory. So, I will not get into the technical analysis of the views of the pioneers of the identity theory in the late 1950s and decide whether they were type physicalist or token physicalist. For this discussion, see John Heil, *Philosophy of Mind: A Contemporary Introduction* (London, New York: Routledge, 1998), pp. 71-85; Kim, pp. 53-62.

¹²⁶ Kim, p. 47.

¹²⁷ *Ibid.*, p. 47.

given situation.¹²⁸ To describe situations of this sort, we sometimes use the expression “his judgment was cloudy”.

Furthermore, any bodily damage will end up with the brain activity wherein a person feels a pain of some sort. Furthermore, under extreme conditions, a subject’s retina may be fooled and, consequently, primary visual cortex can make him feel that he is surrounded by dark forces. At the end of this process, he may naturally form beliefs about dark forces.¹²⁹ Then, as a result, it seems appropriate to conclude that “everything that occurs in mental life has a state of the brain (or the central nervous system) as its *proximate* physical basis”.¹³⁰ More explicitly: “[The mind-brain correlation thesis] For each type M of mental event that occurs to an organism *o*, there exists a brain state of kind B (M’s “neural correlate” or “substrate”) such that M occurs to *o* at time *t* if and only if B occurs to *o* at *t*.”¹³¹

Two points immediately follow from the mind-brain correlation thesis. First, mind-brain correlations are not something that occurs by accident; they are actually “law-like” correlations. The co-occurrence between pain and activation of C-fibers is not a result of strange cosmic accident, but it is the result of lawful regularity.¹³² Second, any kind of change in one’s mental life is supervenient on some specific change in one’s brain state. So, with these two points, identity theorists firmly argue for the claim

¹²⁸ Ibid., p. 47.

¹²⁹ Ibid., p. 48.

¹³⁰ Ibid., p. 48.

¹³¹ Ibid., p. 48.

¹³² Ibid., p. 48.

that “mentality supervenes on brain states and that this supervenience holds as a matter of law”.¹³³

Of course, there are various ways of interpreting mind-body correlations. Descartes, for instance, was one of the philosophers who offered a very radical way of interpreting mind-body correlations with his view of causal interactionism.¹³⁴ Besides Cartesian interactionism, pre-established harmony view, occasionalism, the double-aspect theory, emergentism, epiphenomenalism were all theories that were purported to make sense of those straightforward correlations between mind and body. Of these theories, the mind-brain identity theory of mind was perhaps the most powerful theory of mind, which was launched in the late 1950s with great scientific and theoretical promises. After all, what could have been more promising than the claim that the mind-brain identity theory can formulate a solution for one of the ever-lasting riddles, i.e., the mind-body problem, in the history of philosophy?

According to the identity theory of mind, psychophysical correlations actually indicate a relationship of identity between mental states/events and brain states/processes (more specifically, states in the central nervous system). Just consider the theoretical identities in the natural sciences that “we have discovered from observation and experience”.¹³⁵ For instance, before discovering that a DNA molecule is the same thing as the gene, we had a concept of “gene” that expresses an idea of “an internal factor in the organism that is causally responsible for

¹³³ Ibid., p. 49.

¹³⁴ Ibid., pp. 49-50.

¹³⁵ Ibid., p. 57.

the transmission of heritable characteristics".¹³⁶ So, the gene was something that is described with a reference to "its causal function or role".¹³⁷ After years of empirical research and laboratory work, the molecular biologists finally discovered that it was the DNA molecule that occupies/fills that causal function/role of gene. Therefore, they conclude that the gene is not something over and above the DNA molecule.¹³⁸

The situation cannot be any different in the case of psychophysical correlations.¹³⁹ Just as we come to understand that lightning is not a different phenomenon from the phenomenon of atmospheric electric discharges, we will eventually understand that mental events are not any different from neural processes in the brain.¹⁴⁰ This, of course, does not mean that two expressions have the same meaning. If that were the case, then anyone who knew the meaning of pain, for instance, must have also known the meaning of C-fiber activations. Our ancestors knew a lot about what it is to be in pain, and yet knew nothing about C-fibers. So, the identity claim entails that two expressions, say "pain" and "C-fiber activation", refer to the same phenomenon, but they do not carry out the same meaning.¹⁴¹ To put it simply, there are no two different entities/phenomena picked out by the expressions "mental state" and "brain state"; they are actually one and the same thing.

¹³⁶ Ibid., p. 54.

¹³⁷ Ibid., p. 54.

¹³⁸ Ibid., pp. 54-55.

¹³⁹ J. J. C. Smart, "Sensations and Brain Processes," *Philosophical Review* 68 (1959): 141-156; reprinted in *The Nature of Mind*, ed. David M. Rosenthal (New York: Oxford University Press, 1991), pp. 169-170.

¹⁴⁰ Kim, p. 52, pp. 57-58.

¹⁴¹ Ibid., p. 52.

The psychophysical correlations now boil down to the question about the notion of identity. In order to get clear about the identity statement of mind-brain identity theorists, we must have an idea of what kind of identity-claim they have in mind. Let us say that there are “strict” identities and “non-strict” identities. For an example of non-strict identities, consider our talk when we say that two triangles have the same or equal angles. We do not, of course, mean that these angles are one and the same angle. On the contrary, we mean that they are equal/same only with respect to the magnitude.¹⁴² For an example of strict identities, consider the following case: Thales is the inventor of Thales’ Theorem. The proper name “Thales” and the definite description “the inventor of Thales’ Theorem” happen to pick out, in our world, one and the same individual.¹⁴³ In other words, there is not any property of Thales that the inventor of Thales’ Theorem lacks, or the other way around.¹⁴⁴

Furthermore, identity statements differ epistemically, that is, with respect to the ways they are known. For instance, mathematical truths like “ $7+5=12$ ” and “ $2=\text{the smallest prime number}$ ” are known a priori. In addition to identity statements in mathematics, we have also identity statements often used in natural sciences such as “ $\text{Water}=\text{H}_2\text{O}$ ”, “ $\text{Heat}=\text{molecular motion}$ ”, “ $\text{Light}=\text{electromagnetic radiation}$ ”, and so forth. The truths about these identity statements are based on empirical

¹⁴² Ibid., p. 57.

¹⁴³ Ibid., p. 57.

¹⁴⁴ Ibid., p. 58.

research and observation, so these kinds of identity statements are known a posteriori.¹⁴⁵

Identity theory of mind states the claim that the identity between mental events and brain processes is “strict” and “empirical”. So, when the identity theorist claims that psychophysical correlations are actually signs of the identity between mind and brain, they are committed to the claims (1) that there is not any single property of (a given) mental event that a corresponding brain process would lack or vice versa, and (2) mind-brain identities are empirical in character in which the truth of such identity statements “depends on sophisticated and laborious neurophysiological research”, not on the meanings of mental and physical expressions.¹⁴⁶

Before I go through issues with the identity theory, I would like to make it clear that I am presenting so far type physicalism, not token physicalism. In passing, perhaps it is better to evaluate the difference between them in an outline. Token physicalism does not identify mental states with physical states over their types, but over their tokens. Accordingly, each token of mental state is identified with some token of physical state. This only says that if there is an event or state under consideration and if it has a mental property, then it will also have a physical property.¹⁴⁷ This is, in a sense, a minimalist account of

¹⁴⁵ All examples are from Kim, p. 57.

¹⁴⁶ Ibid., p. 57. One of the readings of the identity claim, that is, each single property of a mental event is also shared by some property of a brain process, leads to some troubles for the identity theorists. See Kim, pp. 62-64.

¹⁴⁷ Kim, pp. 58-62. Also see Ned Block. “Introduction: What is Functionalism?” In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980), pp. 180-181. Token physicalism only says that each particular mental state is a token of some physical state type. Therefore, it allows that

materialism according to which there is a physical description for each occurrence of any mental state; and the correlation between tokens of mental states and tokens of physical states does not need to be systematic at all.¹⁴⁸ Moreover, there is no identity claim involved, say, between mental state types, and brain state types. In contrast, type physicalism asserts the claim that mental state types are nothing over and above physical state types. For each type of mental state, there is a type of brain state, and these types do not refer to two distinct things; they are actually one and the same thing. This is the important difference between type physicalism and token physicalism. Token physicalism is a weak materialist account and it does not entail type physicalism.¹⁴⁹ It is almost a neutral thesis; and it simply says nothing positive or constructive about the supposed connection between mental states and physical states. If you are committed to materialist philosophy of mind, token physicalism does not give you further reasons or evidences to support your materialist worldview.¹⁵⁰ Generally speaking, it is almost standard procedure in philosophy circles to take the mind-brain identity theory as type physicalism, and when people make a comparison between functionalism and identity theory, they all have type physicalism in mind.¹⁵¹ Pain is C-fiber activation. This is the identity statement between

the same mental state can be realized in a wide range of beings with different physical anatomies like that of Martians, homo sapiens, electromagnetically autonomous artificial intelligent systems, etc.

¹⁴⁸ Kim, pp. 58-62; Donald Davidson, *Essays on Actions and Events* (Oxford: Clarendon Press; New York: Oxford University Press, 1980), pp. 207-227.

¹⁴⁹ Kim, pp. 58-59.

¹⁵⁰ *Ibid.*, p. 61.

¹⁵¹ *Ibid.*, p. 60.

two types of states, which are type of a mental state and type of a brain state. Now, let's wrap up the identity theory of mind at the end of our discussion. As far as I conceive, identity theory says something like what follows:

Identity Thesis (IT): A subject, *S*, is in the state of pain *P* at *t*, iff there is C-fiber firing, *C*, in *S* at *t* and $C=P$.¹⁵²

IT stands for the individuation of pain as identified with C-fiber activation. It is obviously an open question whether IT is meant to specify conscious pain or unconscious pain. Perhaps, identity theory of mind aims to reduce consciousness ultimately into physical properties a subject's brain possesses. Then, the question about the difference between conscious and non-conscious pain, for identity theorists, may boil down to the question of finding a natural place for consciousness in microphysical processes. In any event, identity theorists avoid making any reference to intentional and phenomenal properties at the right-hand side of their formulation of mind-brain identity statements, so, they believe, the project of "naturalizing" the mind can be accomplished only in such manner.

Type physicalism was a very strong materialist theory according to which any kind of mental event or process is, in principle, reducible to some kind of brain event or process. For this reason, there was no room for Cartesian souls/ minds in this theory. If mental properties are type-

¹⁵² For a very interesting article about the incorrect use of C-fiber activation in philosophy of mind, see Roland Puccetti, "The Great C-Fiber Myth: A Critical Note," *Philosophy of Science* 44, no. 2 (June 1977), pp. 303-305. Puccetti illustrates the information overload about C-fibers and reveals the common mistake of taking C-fibers as a brain state corresponding to feeling pain. After reading Puccetti's article, I just used the term 'C-fiber activation' only for the purposes of hypothetical illustration of identity theory, not because I believe that C-fibers are what should be taken to be responsible for feeling pain. See also John Heil, *Philosophy of Mind*, p. 78.

identified with physical properties, then, it follows that there could not be any immaterial mental substance.¹⁵³ Behaviorism and type physicalism, in this respect, shares the same goal, that is to exclude Cartesian metaphysics from a naturalistic explanation for the mind; but they differ over the matter whether a mental state is something with a causal power on behavior. For identity theorists, mental state and brain state is one and the same thing, and if brain states have causal impact on behavior, so must have mental states. This is actually regarded as one of the nice outcomes of holding (type) physicalism; because, unlike other theories such as epiphenomenalism and causal interactionism, “it makes mental causation entirely unmysterious: Mental causation turns out to be a species of physical causation”.¹⁵⁴

Everything, as far as type physicalist is concerned, seemed to be smooth and fine. The mind is the brain, full stop, the permanent and comprehensive solution to the chronically unstable relationship between mind/head and body.¹⁵⁵ This is not to say, of course, that the theory was immune to various sorts of criticisms. Of these criticisms, however, one of them is often counted as the fatal blow to the very materialist core of type physicalism. Type physicalism states the identity claim that, for instance, pain is C-fiber activation. But, this has the implication that if a being does not have C-fibers or a brain of appropriate sort, then it will never be able

¹⁵³ Kim, pp. 61-62.

¹⁵⁴ Ibid., p. 56.

¹⁵⁵ For a genius interpretation of the relation between mind and body in a literary form, see Thomas Mann, *Transposed Heads: A legend of India*. Translated by H. T. Lowe-Porter (New York: A. A. Knopf, 1941).

to feel pain.¹⁵⁶ But, it is an empirical fact that some organisms, say reptiles and mollusks, whose brains are radically different from human brains, can feel pain even though they may not have C-fibers in their brains. For these kinds of organisms, having a pain might be nothing over and above D-fiber activation, for instance. In addition to this, it is conceptually possible for inorganic creatures like intelligent extraterrestrials to have a pain even though they differ from organic creatures regarding the biological anatomy. Furthermore, identity theory does seem to damage our common-sensical intuitions for building “intelligent electromechanical systems (that is, robots)”.¹⁵⁷ One of the most passionate goals of Artificial Intelligence research is, beyond any dispute, to build a robot that can feel pain, fall in love, know what it is like to be a robot, etc. The idea of mind-brain identity, however, is a big obstacle to this (conceivable) dream come true. Above all, it would not be possible, within the borders of type physicalism, for the same individual to token different types of brain states (for the same mental state) at different times; yet neurological research provides scientifically accurate reports indicating quite the contrary. For instance, in some cases of brain injuries, the brain reorganizes itself in order to compensate for lost functions in the damaged region of the brain. This is commonly known as “the plasticity of brain”, and identity theory seems to fail to give an account for this phenomenon.

The mind-brain identity theory of mind emerged out of a deep dissatisfaction with the behaviorist framework. Not only does it find a

¹⁵⁶ Kim, p. 69.

¹⁵⁷ Ibid., p. 70.

natural place/seat for the mind, it also revealed a scientific way to understand mind without worrying about calling Cartesian souls back to work. Behaviorists, in a way, were perhaps very much afraid that invoking mental states, understood as internal-causal states of a being/system, would bring the old problems back and this would eventually lead them to accept the idea of “ghost in the machine”. This could have been one of the reasons for their denial of the idea of a mental state producing a behavior. The mind-brain identity theory of mind, on the other hand, clearly explains that these sorts of worries are obviously unwarranted. One can defend the view of a mental state as something causally responsible for the behavior, and identify mental states with brain states, and thereby dismiss all the relevant Cartesian thoughts about the nature of mind. In spite of all these good outcomes, the mind-brain identity theory did not enjoy a long-term authority. Especially with the impact of Putnam’s multiple realizability argument in “Psychological Predicates”,¹⁵⁸ there has been an effectively growing suspicion about the scientific and theoretical value of the mind-brain identity theory, and a new non-reductive approach to mind called “Functionalism” was born in the late 1960s as a replacement for both behaviorism and identity theory of mind.¹⁵⁹

¹⁵⁸ Hilary Putnam, “Psychological Predicates.” In *Art, Mind, and Religion*, edited by W. H. Capitan & D. D. Merrill (Pittsburgh: University of Pittsburgh Press, 1967), reprinted later with the new title “The Nature of Mental States” in *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980a), pp. 223-231.

¹⁵⁹ Kim, p. 70, p. 73.

Functionalism

The very idea of artificial psychology, which implies systems that possess all the genuine and real properties of intelligence and cognition present in human beings and animals except the property of being “natural”, has been the central notion in Artificial Intelligence (shortly as AI) research.¹⁶⁰ AI researchers in computer science, following Turing’s revolutionary heritage, believed that it was both technically and logically possible to build machines or program computers that can realize diverse sorts of cognitive states and activities that include deductive and inductive reasoning, believing, desiring, planning, dreaming, etc., much like the states and activities of which normal adult human beings are capable. For Simon and Newell, building computers that can think was not only logically possible, they believed that soon we will be witnessing that a digital computer will beat the world’s chess champion, it will be able to discover and demonstrate a new mathematical theorem,¹⁶¹ it will compose a fine piece of music, and most importantly with these advances in AI, “most theories in psychology will take the form of computer programs, or of qualitative statements about the characteristics of

¹⁶⁰ B. Jack Copeland, *Artificial Intelligence: A Philosophical Introduction* (Cambridge: Blackwell, 1995), pp. 46-48.

¹⁶¹ Ibid., pp. 7-8. The Logic Theorist, for instance, was one of the most impressive works of AI researchers that succeeded in proving some of the theorems presented in Alfred North Whitehead and Bertrand Russell, *Principia Mathematica* (Cambridge [Eng.]: University Press, 1960). In one case, the proof of the Logic Theorist was found to be more elegant than Russell and Whitehead’s proof. Even Russell admitted the beauty and simplicity of the proof. Newell, Shaw and Simon, so much excited about this achievement, decided to give credit to the Logic Theorist and listed the Logic Theorist as a joint contributor. However, the editor of *The Journal of Symbolic Logic* kindly rejected such a request.

computer programs".¹⁶² This could have been the most robust expression of the initial targets and ideology behind AI. Simon and Newell put forward a firm and strong claim that there is no contradiction in terms about the idea of artificial psychology. In other words, the idea of thinking computers does not involve a conceptual absurdity, unlike the ideas of "four-sided triangle" and "married bachelor". In short, thinking and perceiving, for AI researchers, are more like flying such that both natural and artificial systems can succeed.¹⁶³ After one point, there will be probably no sense to distinguish artificially intelligent systems from naturally intelligent organisms. We will eventually have the same moral obligations toward them, at least so Simon and Newell might believe.

On the other hand, Hilary Putnam, deeply fascinated by the theoretical and technological achievements in AI, began to work out/formulate, in a series of papers in the late 60's,¹⁶⁴ a new conception of mind that incorporates the computer model of mind and the notion of Turing machines to the philosophy of mind.¹⁶⁵ Finally in 1967, Putnam proposed a new conception of mind, later to be called "Functionalism Theory of Mind" (henceforward as FTM), in his famous essay

¹⁶² H.A. Simon and A. Newell, "Heuristic Problem Solving: The Next Advance in Operations Research," *Operations research* 6, no. 1 (1958): pp. 7-8.

¹⁶³ Copeland, *Artificial Intelligence: A Philosophical Introduction*, p. 1.

¹⁶⁴ Hilary Putnam, "Minds and Machines." *Dimensions of Mind*, edited by Sidney Hook (New York: New York University Press, 1960), 148-180, reprinted later in Hilary Putnam, *Mind, Language and Reality* (Cambridge, New York: Cambridge University Press (1975), 362-385; Hilary Putnam, "Psychological Predicates." In *Art, Mind, and Religion*, edited by W. H. Capitan & D. D. Merrill (University of Pittsburgh Press, 1967), reprinted later with the new title "The Nature of Mental States" in *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980a), pp. 223-231.

¹⁶⁵ Cf. Jerry A. Fodor, "The Mind-Body Problem," *Scientific American* 244 (1981), pp. 114-25. Fodor says that "[F]unctionalism that is neither dualist nor materialist has emerged from philosophical reflection on developments in artificial intelligence, computational theory, linguistics, cybernetics and psychology".

“Psychological Predicates” that updates behaviorism and mind-brain identity theory.¹⁶⁶ Since its debut, FTM has been the cornerstone of the contemporary philosophy of mind. This essay achieved three incredible things. First, it played a central role in the decline of type physicalism, the most popular and credible theory of mind back then. As opposed to the chauvinistic attitude in the mind-brain identity theory, it opens up a widely liberal perspective through which any kind of being / system (e.g., computers, souls, extra-terrestrials, and so forth) that can pass the Turing Test is qualified as an intelligent being. Second, it initiated a very revolutionary perspective that changed the way psychology and intelligence are understood and, consequently, it gave rise to FTM. Third, it offered an anti-reductionist view about the nature of mental states.¹⁶⁷ In a nutshell, FTM has brought forth a radical change to our way of describing mentality and intelligence by offering a new conceptual repertoire for the comprehension of the metaphysics of mind, a framework that was not only backed up by largely optimistic AI research predictions but also pumped up with extremely powerful forces of intuition and imagination / conceivability arguments.

There is something extremely incredible and stunningly intuitive in the ideology behind FTM in philosophy of mind. It is my initial aim, in this section, to reveal and examine this ideology from a philosophical perspective. To this purpose, I first present FTM with reference to the

¹⁶⁶ Putnam, “The Nature of Mental States”, pp. 223-231; see also Ned Block, “The computer model of the mind.” In *Readings in philosophy and cognitive science*, edited by Alvin I. Goldman (Cambridge, Mass.: Harvard University Press, 1993), pp. 819-831; Ned Block, “Introduction: What is Functionalism?” in *Readings in the Philosophy of Psychology*, ed. Ned Block (1980), pp. 171-84.

¹⁶⁷ Jaegwon Kim, *Philosophy of Mind* (Westview Press, 1996), p. 73.

basic dichotomies such as “(realized) function vs. (realizing) structure”, “role vs. occupant”, “software vs. hardware” that have been applied by functionalists in establishing their position. Then, I will highlight the similarities and differences between FTM, type physicalism and behaviorism. Afterwards, I will explain the Ramsey-Lewis method for the individuation of functional-causal states of a being/system.

In order to pinpoint the underlying ideology behind FTM, it is crucial to make a proper presentation of the fundamental essence embedded in almost all versions of functionalism. With this broad and proper presentation, I hope, one will be able to come to recognize its thought-provoking and intuition-based reasoning features. Without further ado, let us then begin the evaluation of the essential feature of functionalism. Consider the following simple example of keys. A key can be realized by many diverse physical kinds. For instance, it can be made out of metal, plastic, wood, and so forth. It may also come in many different shapes, weights, colors, etc. But, all these keys have one thing in common, that is to say that their job is to unlock something like door locks, bank vaults, etc. Without this unique role, none of them would be counted as a “key”. Being a key is, at bottom, a functional property.¹⁶⁸

Consider the square function in mathematics; e.g. $f(x)=x^2$. Metaphorically speaking, a function is a black box or a machine which, given appropriate input(s), converts the given input/argument to an output/value. So, in the case of the square function, it is a function such that given the right input(s), say number(s), it will operate on that

¹⁶⁸ A key can be defined in purely functional terms, but a key is not an (probabilistic) autonomous functional system. So, we should keep this difference in mind during the discussion of FTM.

input(s), x , by multiplying it with itself and gives that output/value at the end of the process. It is also crucial to notice that the function always produces the same output for a given input. No matter how many times you try, the square function for a given input, say 2, will always give you the output as 4.

In parallel to this idea of function, let us now think of the question what makes something the kind of being it is. For instance, what does make something a mother of something, the kind of being that instantiates a property of being a mother? In Ursula K. Le Guin's novel *The Left Hand of Darkness*,¹⁶⁹ we have a very different way of understanding this property. Think of two worlds, one is our actual world and the other is a possible world called "Gethen" where people differ from us with respect to their sexual physiology. In our actual world, the notion of being a mother is perhaps defined as follows: being a mother is to be a female individual giving birth to another individual—the essential aspect of being a mother—and taking care of him for his well-being—the social aspect of being a mother. In Gethen, people are androgynies, i.e. having both sexes at the same time, and pick up the role of being a father and being a mother during kimmer period. Now being a mother is not to be identified with being a female individual at Gethen; because the same individual can be both mother and father regardless of his/her sex. The notion of being a mother, as functionalists may claim, must be understood via its functional properties, not with its sexual physiology. So, being a mother is a functional property of any individual such that right input will cause that individual to move to a state of

¹⁶⁹ Ursula K. Le Guin, *The Left Hand of Darkness* (New York: Ace Books, 2000, c1969).

pregnancy and that state will produce a baby (of some kind).¹⁷⁰ Here, the functional property of being a mother does not make any ontological commitment as to determine the sex of a mother.

Consider now this time a functional system like a mousetrap, one that can be said to have some degree of autonomy unlike keys. One can manufacture a mousetrap by making use of different materials, different forms and different methods.¹⁷¹ The essential definition of a mousetrap has nothing to do with which material it is constituted of or which form it must take and by which method it is built up. A mousetrap can be manufactured by using iron, aluminum, gold, etc.; or it can have different forms/shapes that are thought to be suitable for the target; or designers of mousetraps can follow different methods in building up and shaping their mousetraps. All of these things do not matter at all. Any thing X is a mousetrap iff it receives a mousey input; it goes from the state of opened trap into the state of closed trap, and gives the output, i.e. trapped mouse/dead mouse. Therefore, the essence of a mousetrap (i.e., what makes that object the kind of object it is) is its function to produce a trapped/dead mouse whenever it receives a mousey input. A mousetrap is a functional system, and it can be realized in many different ways.¹⁷²

¹⁷⁰ I used Le Guin's example in order to illustrate the basic idea of the functionalist thought. Le Guin's intention with the thought experiment was to show the dichotomy of values behind the inequality between men and women. See Ursula K. Le Guin. "Is Gender Necessary? Redux." In *Dancing at the Edge of the World: Thoughts on Words, Women, Places* (New York: Groove Press, 1989), pp. 7-16.

¹⁷¹ Alternatively, one can also assign a "mousetrap function" to anything that may carry out the power to kill/damage a mousey input. For instance, a portion of poison on a piece of cheese, in this sense, can be interpreted as a mousetrap.

¹⁷² Jerry A. Fodor, "The Mind-Body Problem," *Scientific American* 244(1981): pp. 114-25.

In conceiving the notion of function in this way, FTM gives rise to a well-known notion called Multiple Realizability Thesis (henceforward as MRT), the first premise of a functionalist argument about the nature of mental states.¹⁷³ More explicitly:

1. (MRT) Any type, say ϕ , is variously instantiated =_{def.} It is metaphysically and empirically possible for ϕ to be instantiated by tokens of two or more distinct physical types.¹⁷⁴ [For all types, ϕ , (It is possible_{m/e} that P-tokens instantiate ϕ and Q-tokens instantiate ϕ , and $P \neq Q$).
 2. If mental types are variously instantiated / multiply realizable, then the essences of those types are determined neither by their physical ingredients nor by the physical bonding among those ingredients that happen to be the physical basis of that bonding, but by their functional architecture.
 3. Mental types are variously instantiated / multiply realizable.
- ∴ Functional-causal architecture determines the essence of mental types.

¹⁷³ It is an open question whether FTM entails MRT, or MRT entails FTM. There are some functionalist philosophers who do not agree with the view that MRT is essential component of FTM. Fodor, for instance, does not think that there is a necessary connection between MRT and the notion of "function". From his point of view, one can believe in one but not the other one. See Jerry Fodor, *The Elm and the Expert: Mentalese and its Semantics* (Cambridge, Mass.: MIT Press, 1995), p. 11 and endnote 2 on p. 121. Also on this issue, see David Lewis. "Review of Putnam." In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980), pp. 232-233; David Lewis. "Psychophysical and Theoretical Identifications." In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980a), pp. 207-215; David Lewis, "An Argument for the Identity Theory," *Journal of Philosophy* 63, no. 1 (January, 1966), pp. 17-25. The whole discussion about the question of logical connection between MRT and FTM is nicely presented and examined in Lise Marie Andersen, *Functionalism and Embodied, Embedded Mind – The Extended Story* (M.A. Thesis, Edinburg Research Archive, 2007), pp. 25-35.

¹⁷⁴ Jaworski, William. 24 February 2011. *Mind and Multiple Realizability*. Available [online]: "<http://www.iep.utm.edu/mult-rea/>" [24 February 2011]; Block, "Introduction: What is Functionalism?" pp. 171-84.

With this argument, FTM indicates a very interesting relationship between functional property and structural property. Consider, for instance, a mental property of being in pain. According to functionalists, this mental property is, in essence, a functional property, say F10, such that it is triggered by certain types of inputs, say I10, and tends to cause a certain type of behavior, say O10, and other relevant functional states, say F11 and F12, the latter of which also leads to some other functional state, say F13 and a different kind of behavioral output, say O13. Under this abstract and relational description, there can be numerous physical structures, even spiritual structures, which realize a given functional property such as F10. Modally speaking, it is metaphysically possible, for instance, for a subject to be in pain without having any C-fiber structures in his body. Being in pain for this subject can be realized by the leakage of X-plasma structures. On the other hand, functionalists would strongly hold, there is no possible world where a subject would be in pain and not have a functional property, F10. This is what makes a state of pain the kind of state it is, and hence without this functional property, we would not be capable of feeling any pain whatsoever.¹⁷⁵ Empirically speaking, it is even empirically possible that the same functional property, being in pain, is realized by C-fiber firings in human beings, while it may be completely another story for octopuses to be in pain. Even in the same species, it is empirically possible that distinct physical types in the same individual can realize the same functional property at different periods of his life. In order to envision the difference between function and

¹⁷⁵ But, for some people, one of the essential properties of being in pain is that it hurts. So, it is a matter of debate whether FTM can explain qualitative states.

structure, functionalists mostly apply to the example about the software/hardware dichotomy in which mind is identified with a universal software that runs on radically different sorts of hardware.¹⁷⁶

In a nutshell, FTM states the claim that the intrinsic/essential property of a mental state has to do with its causal role in a network of functional architecture of a being/system, and it has nothing to do with what (kind of structure) realizes or occupies that role. Hence the distinction between role and occupant of mental states. What occupies the role of a mental state, say pain, may be C-fiber activation in humans and X-plasma leakage in extraterrestrials, but the functional role of a mental state is independent of the (the physical/spiritual) structures that happen to occupy that role. FTM simply asserts the idea that electric circuits, neurological states, soul-stuff, etc., can be the physical/spiritual structures that realize the same functional state in different kinds of beings/systems as long as these different kinds of beings/systems instantiate functionally equivalent systems.¹⁷⁷ More explicitly and generally, FTM states the following claim about the nature of mental states:

¹⁷⁶ Ned Block, "The computer model of the mind." In *Readings in philosophy and cognitive science*, edited by Alvin I. Goldman (Cambridge, Mass.: Harvard University Press, 1993), pp. 819-831; Kim, *Philosophy of Mind*, pp. 75-76, p. 91.

¹⁷⁷ Of course, it is not enough to claim that mental states are variously realizable in different physical bodies/soul-stuffs; one needs to support this claim in order to demonstrate the validity of the overall argument. In order to back up MRT, FTM takes two lines of support, one is in conceptual and the other is in empirical form. I am afraid that the discussion of these two topics as well as negative insights against them will lead me to a very long-debated and controversial issue, and since my purpose is to draw FTM in general lines, I will not discuss the technical details about MRT. There is a huge literature about MRT. For a nice start, see Chris Eliasmith, "The Myth of the Turing Machine: The Failings of Functionalism and Related Theses," *Journal of Experimental and Theoretical Artificial Intelligence* 14, no. 1 (February 2002), pp. 1-8; John Heil, "Multiple Realizability," *American Philosophical Quarterly* 36, no. 3 (July 1999), pp. 189-208.

Any being/system B is in a type M of mental state, iff there is a state of B that plays a unique functional/causal role F definitive of M-type within a complex network of states of B such that it mediates, together with other relevant states in that network of states, between perceptual inputs and behavioral outputs of B.¹⁷⁸

It is a purely empty and futile enterprise, according to the functionalist metaphysics of mind, to explore the source of psychological laws, which govern mental operations and behaviors of material or spiritual beings, in the underlying (material/spiritual) structure/substance where those laws make their first appearance. This is another way of saying that there is no difficulty, in principle, for the individuation of the same mental states in physically/spiritually differently composed beings and substances, provided that those beings and substances manifest functionally equivalent architectures, i.e., a functional organization of an appropriate kind that is constitutive of mentality and intelligence. As thus construed, functionalism surprisingly makes both materialists and spiritualists happy, probably the first time in the history of mankind.

Materialism and Spiritualism are two doctrines in the history of humankind, doctrines that offer completely distinct comprehensions of the universe. It is very unlikely, for most philosophers, to witness that there will be a point of agreement between them any time soon. To be a witness of a new theory that could meet the expectations of both materialists and spiritualists could be regarded as a mere fantasy or crazy dream of a lunatic man. Contrary to this pessimistic outlook, FTM seems

¹⁷⁸ This formulation is mostly inspired from Armstrong's definition for the concept of mental state and Block's exposition of functionalism. See David Armstrong. "The Causal Theory of Mind." In *The Nature of Mind*, edited by David M. Rosenthal (New York: Oxford University Press, c1991), p. 183; Ned Block. "Introduction: What is Functionalism?" In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980).

to present a new way of looking at mind that can literally put a happy face on both materialists and spiritualists.

Let us consider pain again. Since FTM leaves it open to individuate pain types for various kinds of beings, it is possible to formulate what it is to be in pain for both material and spiritual beings. For instance, on the materialist construal of pain, the following can be taken as the description of pain for material beings:

A material subject, M, is in pain, iff there is a functional state, say F10, such that it mediates, together with the other functional states, between bodily damage and avoidance / escape behavior from the source of that bodily damage.

We can construct a similar description of pain for spiritual beings:

A spiritual subject, S, is in pain, iff there is a functional state, say F10, such that it mediates, together with the other functional states, between spiritual damage (as the effect of psychological disturbance of another soul) and avoidance / escape behavior from the source of that spiritual damage.¹⁷⁹

Any mental state, intentional or phenomenal, can be realized in material or spiritual substances / structures, and it just does not matter at all in which kind of substance it is *embodied* or *ensouled*. The material or spiritual host of mental states does not contribute to the essence of those states; they are required only for the realization of those states. In a sense, they help them come into being, but they do not determine the essence of mental states.

After having discussed the fundamental aspects of functionalism, let us now go back to behaviorism and identity theory, and highlight the

¹⁷⁹ In general, most functionalists today, excluding a very few philosophers, are materialists. With these two formulations above, I particularly want to emphasize, contrary to the received impression, functionalism is a neutral thesis about the substance / structure of mental states. In other words, functionalism does not give a further reason to take a side with materialism or spiritualism.

similarities and differences between FTM and those two theories of mind. Let us begin with the contrast between the mind-brain identity theory and FTM. Mind-brain identity theory and FTM actually both take mental states as inner states of systems/beings with causal powers on producing behavior, but they differ over the matter whether those inner states should be type-identified with neurobiological states of mental beings or more abstract and formal properties that beings/systems instantiate during mental events and processes. For a type physicalist, the essence of mental states is identified with brain state types. For instance, having a pain is nothing over and above the activation of C-fibers. A functionalist, on the same point, rejects the definition of type physicalist and he claims that having a pain is to be in a functional state, together with other relevant inner states, that mediate between perceptual inputs and behavioral outputs of mental beings. So, according to FTM, what makes any given mental state the kind of state it is is to perform a certain job, X-ing, not its having a neurobiological structure or any other physical/spiritual structures.¹⁸⁰

As for the contrast between behaviorism and FTM, we can say that both of them consider “sensory input and behavioral output—or ‘stimulus’ and ‘response’— as central to the concept of mentality”.¹⁸¹ From this angle, one can interpret FTM as the view that keeps the broadly behaviorist criterion intact. But, it is also important to keep in mind that FTM is not a kind of behaviorism, because while a behaviorist does not believe that a mental state is the cause of behavior, a functionalist is

¹⁸⁰ Kim, *Philosophy of Mind*, p. 77.

¹⁸¹ Kim, p. 77.

strongly committed to the claim that the defining aspect of a mental state is its functional/ causal role in producing a behavior.¹⁸²

As to the difference between behaviorism and FTM, mental states, in functionalism, are identified with “real internal states” of a subject “with causal powers”, whereas mental states, in behaviorism, are identified with complex dispositions that exhibit a certain kind of behavioral patterns.¹⁸³ According to the instrumentalist approach of behaviorism, dispositions for a behaviorist only mean a conditional or hypothetical property. On the contrary, dispositions for a functionalist indicate real states of a subject that explain the causal correlation between a given disposition and its behavioral outcome. For instance, a behaviorist explains the property of water solubility as the disposition for a certain conditional or hypothetical statement to be true of it. If we have two objects, one that dissolves in water while the other does not, a behaviorist would say that the conditional statement “if placed in water, then it will dissolve” is true for the former but not for the latter. It is a brute fact, for a behaviorist, that the conditional statement is true for the first but not for the latter object.¹⁸⁴ However, for a functionalist, there is a different story. It is due to the fact that the former but not the latter has a real state or property that the former object, say a sugar cube, tends to dissolve in water. In this context, functionalism takes a realist attitude and gives a causal explanation for the water solubility of the object.¹⁸⁵ So,

¹⁸² Ibid., p. 77; Block, “Introduction: What is Functionalism?” p. 176.

¹⁸³ Kim, pp. 77-78; Block, “Introduction: What is Functionalism?” pp. 175-76.

¹⁸⁴ Kim, *Philosophy of Mind*, p. 78.

¹⁸⁵ Ibid., p. 78.

for behaviorism, mental states are like fictional or instrumental parameters in order to give an account of the behavior of organisms. The real entities for a behaviorist only consist of observable physical stimulus conditions and observable behavioral patterns associated with those conditions, whereas, for a functionalist, “mental states are inner causes of behavior, and as such they are in addition to behavior”.¹⁸⁶ FTM is not a weak claim that for any given mental state there exists an appropriate functional description or model by means of which that mental state is understood in functional terms. For FTM, mental states are functional kinds. That is to say, any mental state is, at bottom, a functional state (of appropriate kind).¹⁸⁷

As a second difference, it is generally thought that behaviorism can't give an account for the nature of mental states, because, unlike the behaviorist view, any mental state is related to other mental states and conditions. According to a behaviorist, a subject's, say S's, belief that p can be translated via his verbal behavior such that, when asked “Is it the case that p?” he would answer “Yes, it is the case that p”.¹⁸⁸ Even here,

¹⁸⁶ Ibid., p. 79.

¹⁸⁷ Block, “Introduction: What is Functionalism?” p. 172; Kim, *Philosophy of Mind*, p. 91. I have a reason for adding the specification “of appropriate kind” in a parenthesis. Functionalism, understood very broadly, may give the impression that any kind of functional organization, regardless of the degree of its complexity or primitivity, is a sign for intelligence and psychology. But, mousetraps and coke machines, being very primitive functional organizations, do not have a psychology on their own. Obviously not all kinds of functional organization can be counted as realizing psychological states. For this reason, I suspect whether early frontiers of FTM, including Putnam and Fodor, hold such a bold and implausible claim. The real trouble, on the other hand, for functionalists is to unpack the expression “functional organization of appropriate kind”. This may prove to be the most difficult task for functionalists. I tend to believe that without specification of functionally appropriate kinds, functionalism will always remain a very general and yet uninformative theory of mind.

¹⁸⁸ Kim, *Philosophy of Mind*, p. 32.

one can start to see that behaviorism is in need of adding extra clauses such as “if S wants to tell the truth”. If this is the case, then a behaviorist needs to add the clause of the “desire” to tell the truth and this seems to threaten the behaviorist project.¹⁸⁹ Furthermore, besides other relevant mental states, one needs to specify mental conditions, like “being alert”, “not having any other activity that can be disturbing” and so forth, in order to give a full characterization for the conditions of a subject uttering the sentence “Yes, it is the case that p”. Functionalism seems, for many people, to be an improvement on the behaviorist framework on the issue of the complexity of mental states. Instead of describing any single mental state only in terms of a set of input-output clauses associated with that mental state, “the functionalist will include reference to other mental states in the characterization of a given mental state”.¹⁹⁰ For FTM, in addition to input-output clauses, each mental state must be identified within the causal-functional system of which it is a part.

Let all this suffice for the task of distinguishing FTM from type physicalism and behaviorism. Now, let us proceed to examine one of the important troubles of functionalism. Functionalist philosophers argue for the claim that any state is a mental state only if that state has a particular role in a functional organization in which it mediates, together with the

¹⁸⁹ Kim, pp. 33-34. Contrary to Kim’s conviction, I do not believe that this is a big problem for a good behaviorist, because any good behaviorist can say each behavior is the result of family of dispositions, not only one single disposition. For instance, in the case of the extra clause “S wants to tell the truth”, a good behaviorist can explain this clause as S’s disposition to be honest. “Honesty”, for a behaviorist, is only one kind of behavioral traits, and “being honest” can be explained as gaining a certain kind of behavioral disposition to be a responsible member of society. So, it is possible to give a behavioral characterization of the desire to tell the truth. As thus construed, I can say that there is actually no difference between behaviorism and functionalism. Both can explain mental states in terms of their relation to other mental states. The real threat for both behaviorism and functionalism can come from the problem of circularity.

¹⁹⁰ Ibid., p. 79.

other states in that organization, between perceptual stimuli and behavioral responses. This is a good way of describing the nature of mental states; because each mental state is distinguished from the other mental state with respect to its particular functional role between perceptual inputs and behavioral outputs associated with that mental state. Consider the following example. Pains are typically issued from the cause of a laceration of the ligaments of a joint, and cause the emission of high-pitch sound plus tears from eyes, and also cause some other states, one of which in turn causes a facial expression of stress. Itches, on the other hand, are typically issued from skin irritation and causes scratching and some other mental states such as the desire to scratch constantly which in turn leads to the behavior that relieves that constant desire. But, in order to have pain or itch in the first place, there must be some other requirements that functionalists must meet. First, we need to take the fact that a subject must have a normal nervous system into account. Again, we need to talk about some other mental states that make it possible for a subject to be in pain such as being alert for the task. Furthermore, while explaining what it is to have pain or itches, functionalists employ some other mental states like anger, desire to get rid of pain, etc. In the final analysis, it seems that either this process of making reference to other mental states will go on ad infinitum or trap functionalists in a vicious circle.¹⁹¹ Lewis claims that instead of the effort to individuate mental states one by one and bite the bullet of circularity, with the help of Ramsey sentences, one can provide a more plausible way to understand

¹⁹¹ Ibid., pp. 104-05. The way a machine functionalism and non-machine functionalism handles the issue is different. See Kim, chapter 4 and chapter 5 for a comparison.

functionalism through which mental states are to be individuated once and for all. This is a holistic approach for individuating mental states.

David Lewis, in his essay "Psychophysical and Theoretical Identifications",¹⁹² attained success in resolving the question of circularity addressed to all versions of functionalism. In all versions of functionalism, a mental state is characterized with respect to its intermediary causal role between inputs and outputs, where inputs can be perceptual stimuli and other mental states as causes and outputs can be behavioral outputs and other mental states as effects. Let us have a look at Lewis' way of describing functionalist characterization of mental states.¹⁹³ Suppose that functionalism is a psychological theory, T, purported to explain the relation of a mental state to other mental states, sensory inputs and behavioral responses of a subject. Let us leave out the relation of a mental state to sensory inputs and behavioral outputs in our analysis just for the sake of simplicity and focus only on the relation of a mental state with other mental states of a subject. Let us also suppose that the terms standing for psychological elements in T, i.e., mental states, are going to be singular terms. So, instead of saying that a subject, A, is angry, now we can equivalently say in T that there are states, s₁ ... s_n and A has s₁₀=anger. More explicitly:

T [(s₁ ... s_n) and A has s₁₀]

¹⁹² David Lewis. "Psychophysical and Theoretical Identifications." In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980), pp. 207-215.

¹⁹³ I will follow Block's exposition of Lewis' way of explaining mental states; because it provides a much more simplified version of Lewis' method. Cf. Block, "Introduction: What is Functionalism?" pp. 174-75.

By applying Ramsey's method, now let us replace each occurrence of s in T with correspondent variables as " $x_1 \dots x_n$ ". Also, let us put an existential quantifier in front of the entire sentence in order to quantify over variables. Now, we have the following Ramsified sentence:

$\exists x_1 \dots \exists x_n [T(x_1 \dots x_n) \text{ and } A \text{ has } x_{10}]$

Now, functionalists contend that one can give a functional description of what it is to have pain:

A subject, A , has pain, if $\exists x_1 \dots \exists x_n [T(x_1 \dots x_n) \text{ and } A \text{ has } x_{11}]$

Let us illustrate how the formulation above works. Suppose, again for the simplicity, that pain is something caused by a laceration of the ligaments of a joint and it typically causes scream-cry behavior and a mental state of desire to get rid of the pain from the body and a mental state of anger, and anger in turn causes the production of a stressed facial expression. Accordingly then, assuming that this is the common point that all pain-feeling organisms have, the Ramsey sentence of this state of pain will be something like the following:

$\exists x_{11} \exists x_2 \exists x_3 [(x_{11} \text{ is caused by a laceration of the ligaments of a joint and } x_{11} \text{ causes } x_9 \text{ and } x_{10} \text{ and emission of a high-pitch sound and tears from eyes, \& } x_{10} \text{ in turn causes a stressed facial expression) \& } A \text{ has } x_{11}]$

Of course, if we begin to merge input and output clauses with this Ramsey sentence of pain, then we will have a much more complex picture about what it is to have pain. But, leaving aside the whole functional analysis of pain along with input and output clauses, the formula above nicely illustrates a very decisive insight of FTM. In conceiving pain in the way illustrated above, FTM maintains that pain is identified as a real state that any subject has when that state is caused by

a laceration of the ligaments of a joint, and causes the emission of high-pitch sound plus tears from eyes, and also causes some other states, one of which in turn causes a facial expression of stress.

The upshot of all this discussion is that there is no distinction between the functional/abstract description of terms like “carburetor” and “ignition chamber”, and mental state terms like “pain” and “anger”. By making use of the same method, it is also possible to give a definition of some item to instantiate the property of being a carburetor or an ignition chamber. Of course, we just restrict our definition of pain with a very simple set of items related to pain. If mental states like pain, anger, desire to get rid of pain, etc., are much more complex than the simple functional description above, then functionalism, at least, shows how to get around this complexity¹⁹⁴ with the way it identifies mental states in a Ramsey sentence. Pain is a certain type of functional state such that it is deeply connected to other functional states (composing together a functional architecture) and it is also deeply connected to “the real world only via its relations, direct or indirect, to inputs and outputs”.¹⁹⁵

To sum up, FTM offered a very radical perspective through which mentality and intelligence is understood on a very abstract, general and

¹⁹⁴ It is one of our philosophical practices to take some of the concepts, such as complexity above, to be something like everybody has a clear and distinct idea of what complexity means, and build our reasoning on those supposedly well-known concepts. It is much more plausible, of course, to give a definition of complexity. This is beyond the borders of this thesis project. For a very interesting and fruitful discussion of complexity from a biological point of view, see Richard Dawkins, *The Blind Watchmaker* (New York: Norton, 1986), chapter 1. For a recent philosophical discussion of complexity with a special reference to intentionality, see Peter Godfrey-Smith, *Complexity and the Function of Mind in Nature* (Cambridge; New York: Cambridge University Press, c1998). Godfrey-Smith describes representational capacities over the adaptive skills organisms acquire in response to environmental parameters. For a criticism of this view, see Kim Sterelny, “Basic Minds,” *Philosophical Perspectives* 9, AI, Connectionism and Philosophical Psychology (1995), p. 253.

¹⁹⁵ Block, “Introduction: What is Functionalism?” p. 175.

realistic level, the level that type physicalism and behaviorism were unable to reach. There is something extremely incredible and stunningly intuitive in the ideology of this perspective. First of all, it incorporates the computer model of mind in which mind is understood as a universal software that runs across vastly different physical/spiritual hardware/structures. In this respect, it provides a philosophical ground for the (conceivable) dream of AI come true. Secondly, it was always regarded as the theory that can speak to one's common sense.¹⁹⁶ In this regard, it is like a magical theory that can have equal impact on (and be useful for) every intellectual creature. If one's theory of mind, in a sense, is a tool that helps the individual to understand the mentality and behavior of others and consequently helps that individual to communicate and exchange ideas and thoughts with others, and if functionalism can have, as a theory of mind, equal impact on every intellectual creature, then it follows that even extraterrestrials would apply the same functionalist framework in order to model our animal psychology. This is, in any sense of the word, simply amazing.

¹⁹⁶ Putnam, later on in his career, has become one of the most passionate and strong enemies of functionalism. It is quite ironical, as a matter of fact, that when the founder of FTM has changed his mind, the obedient practitioners of the ideology of FTM have charged him of being a betrayer and weak character. See Hilary Putnam, *Representation and Reality* (Cambridge, Mass.: MIT Press, 1991, c1988), Introduction xi-xii; Hilary Putnam, "Putnam, Hilary." In *A Companion to the Philosophy of Mind*, edited by Samuel Guttenplan (Cambridge: Blackwell, 1994), pp. 507-513; Hilary Putnam, "Functionalism: Cognitive Science or Science Fiction?" In *The Future of the Cognitive Revolution*, edited by David Martel Johnson and Christina E. Erneling (New York, Oxford: Oxford University Press, 1997), pp. 32-44. For a nice survey and discussion of Putnam's change of mind about functionalism, see Oron Shagrir, "The Rise and Fall of Computational Functionalism." In *Hilary Putnam (Contemporary Philosophy in Focus)*, edited by Yemima Ben-Menahem (Cambridge, New York: Cambridge University Press, 2005), pp. 220-250.

The Platonic Ideology

I tend to believe that there exists a hidden Platonic ideology behind functionalism, operative in the background of almost each version of it. This ideology, a two-leveled explanatory framework, is often manifested with its constant emphasis on the software/hardware dichotomy. A few words about Plato's metaphysics are in order.

In Plato's metaphysics, Ideas/Forms stand for the perfect, eternal, never-changing paradigms or archetypes that exist in the realm of Being; whereas objects/things in the realm of Becoming are imperfect and subject to change over time. Platonic Forms are, in a sense, ideal, (ultimately) real and perfect models in which particular things in the realm of Becoming share or participate. In other words, those ordinary objects are only exemplifications or realizations of Forms. Logically speaking, the very existence of ordinary objects (logically and ontologically) depends on the existence of Forms, because without any Form that exists to be shared, there could not be ordinary objects too.¹⁹⁷ Thus, the Form X exists independently of X-ing objects, but X-ing objects cannot exist without the Form X. Metaphorically speaking, objects in the material world are like shadows of those ultimately real Forms in the world of Being. Within this context, a soul, according to Plato, is a miserable and unhappy prisoner in a body with forgotten memories, and its freedom and capacity to know the real nature of things lies in its own intellectual revolution by means of which it comes to a gradual

¹⁹⁷ Richard A. Watson, *Representational ideas: From Plato to Patricia Churchland* (Dordrecht; Boston: Kluwer Academic Publishers, 1995), pp. 5-7.

recognition of the fact that ultimate reality lies beyond the shadows in the material world.

It is striking to notice that FTM is deeply rooted in a Platonic metaphysics. In much the same spirit, FTM asserts the claim that the function of something (e.g. keys, coke-machines, computers) is independent of the physical aspects of what realizes it. Then, it follows, according to functionalists, that a mental state must be identified with a functional state that is independent of the physical/spiritual make-up of whatever *embodies* or *ensouls* it. In functionalism, there is a two-leveled conceptual scheme according to which any given mental state is described with reference to its physical/spiritual composition or its (potential or actual) causal role in a vast array of functional architecture. From a functionalist point of view, lower-level aspects of mental states (e.g., neural firings, synapses in the neural networks) only indicate what mental states are actually made of with respect to one kind of intelligent beings (e.g., human beings and some animals), but they do not explain what constitutes/makes up mental states in general. A pain, for instance, can be realized by C-fiber activations in human beings, or it can be realized by D-fiber activations in octopuses, or, even, it can be realized by some kind of plasma in extra-terrestrial creatures. There must be a universal pain in virtue of which (physically/spiritually) different kinds of pain-suffering beings/systems are said to be in pain. To explain this conceivability of a universal psychology, FTM provides a very liberal account of mind according to which mental states are, in essence, functional states. Functional states, in turn, stand for higher-level properties of intelligent beings, intrinsic properties that give something a

definitive character, e.g. being a kind of intelligent being/system. For FTM, it is not the fine structural details about the material/spiritual host, but the functional details about the complex causal network that determine whether a being/system has a share in universal mentality and intelligence. Functionalism, therefore, rests on the assumption that there is a universal psychology that can be shared across physically/spiritually diverse beings/systems. This ideology is often illustrated with the software/hardware dichotomy. Heil, during his description of functionalism, says that:

Every program is “embodied,” perhaps, in some material device or other. But the very same program can run on very different sorts of material device. In the same vein, we might suppose that every mind has some material embodiment, although minds may have very different kinds of material embodiment. In the case of human beings, our brains constitute the hardware on which our mental software runs. Alpha Centaurians, in contrast, might share our psychology, our mental software, yet have very different, perhaps non-carbon-based, hardware.¹⁹⁸

This is one of the fine instances of the computer model of mind according to which mind is nothing but a universal software that can run on radically different sorts of physical/spiritual stuff. Intelligent beings are intelligent, because all of them share the same universal psychology/software regardless of the differing (material or spiritual) ways they realize that psychology/software. This way of characterizing mind, however, gives rise to a number of questions. As far as I can see, there could be two groups of questions, the former of which is related to the identity conditions of a software and the latter of which is related to

¹⁹⁸ John Heil, *Philosophy of mind: A Contemporary Introduction* (London; New York: Routledge, 1998), p. 91.

the emergent need to clarify notions of “software”, “hardware” and “computer”. So, let us begin with the first group of questions.

Can there really be a universal software/program that can run on different kinds of hardware? The very idea of universal software implies the claim that it is possible to run exactly the same program in physically diverse machines/hardware. But, the following question must be answered first: Under which conditions is it true to say that X is exactly the same program running on different machines? From my point of view, the idea of universal software must be, at least, a matter of debate, yet most people find it obvious and straightforward. Consider the following case. Suppose that you have a program running so smoothly on one of the most advanced computer hardware technologies available now, a computer that is composed of high-tech silicon chips and electronic circuits. Just think what would happen if you want to use exactly the same program in old computers composed of valves and gears. Can it be installed in the first place? One can plausibly argue for the claim that the way to program old computers must be extremely different from the way to program recent computers, because their computational architecture is radically different due to the physical structure of their components. Installing the same program, however, might not be the real problem after all; one can find a good medium to do the work of installation. But, one can go on to insist that installing the same program will require much labor to modify the algorithm in order to adapt computational and physical capacities of the old hardware. So, if the algorithm is the heart of any software, then changing the structure of an algorithm is nothing but playing with the identity conditions of

software. If this is true, then we can no longer talk about the identity of software over physically diverse machines.¹⁹⁹

As for the second set of questions, it is not really clear whether software is something entirely independent of hardware. In general, functionalists seem to hold a view that software-properties of X mark a completely different and independent category than those hardware-properties of X. Does this distinction between software-properties of X and hardware-properties of X really entail that software is entirely independent of hardware? Or, broadly speaking, does the distinction between pure functional properties and pure structural (physical/spiritual) properties entail that function is independent of matter? Let's have a look at Moor's warning about the misconception of "software" and "hardware". Moor²⁰⁰ points out that:

Computer programs are often considered to be part of the software of a computer as opposed to the hardware. Unfortunately, computer hardware is frequently characterised as 'the physical units making up a computer system' (Chandor [1970], p. 179). By contrast this seems to suggest that software is not part of the computer system or even worse that it is not physical. It is important to remember that computer programs can be understood on the physical level as well as the symbolic level. The programming of early digital computers was commonly done by plugging in wires and throwing switches. Some analogue computers are still programmed in this way. The resulting programs are clearly as physical and as much a part of the computer system as any other part. Today digital machines usually store a program internally to speed up the execution of the program. A program in such a form is certainly physical and part of the computer system. Furthermore, since programming can occur on many levels, it

¹⁹⁹ Chris Eliasmith, "The Myth of the Turing Machine: The Failings of Functionalism and Related Theses," *Journal of Experimental and Theoretical Artificial Intelligence* 14, no. 1 (February 2002), pp. 1-8.

²⁰⁰ James H. Moor, "Three Myths of Computer Science," *The British Journal for the Philosophy of Science* 29, no. 3 (September, 1978), pp. 213-222.

is useful to understand the software/hardware dichotomy as a pragmatic distinction.²⁰¹

The whole point boils down to the issue whether software-properties and hardware-properties of a computer indicate a metaphysical distinction or a pragmatic distinction. Moor has a way to settle this issue. For him, one can describe any given computer in terms of its physical aspects or symbolic aspects. For instance, a computer, in a physical description, is an entity composed of silicon chips, electronic circuits and so forth. If you hold on to this description and know enough about the laws of physics about electronics, Moor suggests, you will be able to predict any computer behavior. In addition, one can describe any given computer with respect to its symbolic features. Consider a chess-playing computer. In this case, a computer is a system that can take certain kinds of inputs as symbols, go to a set of states relevant to that input, make use of its database and algorithms, and choose the best output in accordance with inputs. If you detect, for instance, that the computer always moves its queen too early, and know enough about the logical structure of the computer, then you may easily spot the source of the problem.²⁰² By following this strategy, for Moor, one can also provide two different ways to characterize a computer program, the former of which might emphasize its physical properties such as “series of punched cards, configurations on a magnetic tape”, while the latter might take its

²⁰¹ Ibid., p. 215.

²⁰² Daniel C. Dennett, “A Cure for the Common Code?” In *Brainstorms: Philosophical essays on mind and psychology* (Cambridge, Mass.: MIT Press, 1978), p. 107. The example belongs to Dennett.

symbolic properties such as “a set of symbolic instructions”.²⁰³ In the final analysis, Moor wanted to show that software and hardware terms do not actually signify mutually exclusive categories. Just because one can explain software properties of a computer without a reference to its physical aspects, this does not, in any sensible way, prove that software properties are independent from hardware properties. The software/hardware distinction is a pragmatic distinction, and programming, as Moor and Lycan made painstaking efforts to show in detail, is relative to one’s perspective.²⁰⁴

Instead of explaining Form as something connected with Matter, Platonic metaphysics takes it to have a higher-order mode of being, one that is independent of ordinary material objects and the

²⁰³ Moor, p. 213.

²⁰⁴ Moor, pp. 213-222; William G. Lycan, “The continuity of levels of nature.” In *Mind and cognition: A reader*, edited by William G. Lycan (Cambridge, Mass., USA: Basil Blackwell, 1990), pp. 77-96. See also Peter Suber, “What is Software?” *The Journal of Speculative Philosophy* 2, No. 2 (1988), pp. 89-119. Suber believes that we must begin a metaphysical investigation in order to understand what kind of thing software is, and he comes to the conclusion that software is hardware. So, while Moor and Lycan believes that the distinction between software and hardware should be understood as a pragmatic distinction, Suber goes further by claiming that software is hardware. For a recent metaphysical work on the nature of software and criticism of Suber’s and Moor’s views, see Nurmey Irmak (forthcoming), “Software is an Abstract Artifact,” *Grazer Philosophische Studien*. There is not much of a work on the metaphysics of software, and I believe that Irmak has provided the most fruitful account so far. In a personal discussion, Irmak said that his account actually implies the fact that the very idea of taking software as the same thing as mind suffers from the misconception about the nature of software. Software is something produced; therefore analogy/identity between software and mind breaks down from the very beginning. For an extensive discussion of the points concerning software/hardware dichotomy within the context of artificial intelligence and functionalism, see especially Stephen M. Kosslyn & Gary Hatfield, “Representation Without Symbol Systems,” *Social Research* 51, no. 4 (Winter 1984), pp. 1031-1035. The article especially draws attention to the symbol system hypothesis in artificial intelligence literature and functionalism in philosophy of mind literature, and illuminates the weak assumptions grounding both. For instance, among other things, people believe that differences in capacity and efficiency in computers do not matter, and that’s why one can write the same program for two different computers having different storage capacities and CPU speeds. But, this is an extremely wrong way of thinking; because these factors “make a big difference in which programs can actually be run on a machine in real time”, p. 1028. On a similar topic, also see Chris Eliasmith, “The Myth of the Turing Machine: The Failings of Functionalism and Related Theses,” *Journal of Experimental and Theoretical Artificial Intelligence* 14, no. 1 (February 2002), pp. 1-8.

logical/ontological ground of them. Likewise, FTM takes functional properties of mental states as higher-order properties that are independent of physical/spiritual properties of their realizers/hosts and the existence of those functional properties in any body/soul actually explains the reason of that body/soul being intelligent and conscious. In particular, FTM accepts a very Platonic idea that there is universal software that can run on and be shared across physically/spiritually diverse kinds of beings/systems. At the end of this section, I provided a number of reasons why this way of characterizing mentality and intelligence is very problematic. Alternatively, one can adopt an Aristotelian metaphysics according to which the Form of something is intimately connected with the Matter of that being/system. In the context of our discussion of functionalism, one can plausibly argue for the claim that functional properties of mental states are not independent of the physical/spiritual properties of their realizers/hosts; they actually emerge directly out of the physical/spiritual properties of their hosts.²⁰⁵ So, there is an intimate connection between the role X and the (physical/spiritual) occupant of that role X.²⁰⁶ This way of reading functionalism, I believe, may prove to be much more in line with common sense and can be more helpful in dealing with further issues

²⁰⁵ Cf. Stephen M. Kosslyn & Gary Hatfield, "Representation Without Symbol Systems," *Social Research* 51, no. 4 (Winter, 1984), p. 1027.

²⁰⁶ Cf. David Armstrong, *A Materialistic Theory of the Mind* (London: RKP, 1968); David Lewis, "An Argument for the Identity Theory," *Journal of Philosophy* 63, no. 1 (January, 1966), pp. 17–25. Also see V. S. Ramachandran & Sandra Blakeslee, *Phantoms in the Brain: probing the mysteries of the human mind* (New York: William Morrow and Company, Inc., 1999), p. 264. The authors remark here: "Yet a vast majority of psychologists—called functionalists—cling to the view that we can understand mental processes from a strictly computational, behaviorist or "reverse engineering perspective"—without bothering with the messy stuff in the head. When dealing with biological systems, understanding structure is crucial to understanding function—a view that is completely antithetical to the functionalist or black box approach to brain function."

related with artificial psychology, mental representation and intentionality.

CHAPTER IV

FROM CLASSICAL FUNCTIONALISM TO POST-FUNCTIONALISM

The main purpose of this chapter is to introduce the transition from classical functionalist programmes to post-functionalist programmes like the one of Dretske and Millikan's theories. In order to understand Dretske and Millikan's theories of representation, we must have an idea about how this transition takes place. In the beginning of this chapter, I will take you to a brief historical tour into the philosophical climate of the twentieth century philosophy of mind. I will especially draw attention on the connection between intentionality and representationality. Then, I will talk about two different ways of understanding the concept of representation in the philosophy of mind. In the end, I will turn back to classical functionalism, one that has been advertised by Putnam in the late 1960s. Functionalism was not only supposed to bring forth a new metaphysics of mind. For Putnam, it was supposed to give an account for all kinds of intentional/representational states. After Putnam's essay "The Meaning of 'Meaning'", something has been changed in this climate and post-functionalist programmes began to make their first appearances on the stage between the 1980s and 1990s.

Intentionality and Representation

So much has been said about intentionality and representation in the history of philosophy. Especially between the 1980s and 1990s in the literature, philosophy of mind has taken a very powerful and incredible

spin towards the route to explain intentionality and mental representation by means of a functionalist-naturalist framework. At those times, philosophers were not much interested in consciousness as recent philosophers are nowadays. The central idea was that a subject of mind could be intentional without being conscious.²⁰⁷ Something happened, and this idea was replaced by another idea, which is that having intentional states go through being conscious of them. Hence, as far as I understand, the whole focus turned from intentionality to consciousness in contemporary philosophy of mind. Now, the big question is to resolve the mystery of consciousness: “how can our conscious experience fit into the growing body of scientific knowledge about the mind and the brain?”²⁰⁸

As for the era between the 1980s and 1990s, however, the big question was the mystery of intentionality: how can our thoughts reach out and be about other things in the world?²⁰⁹ In particular, how can material states of mind, differently from other material states in nature, bear content/representation? Here, let’s pause for a minute and give a short description of intentionality.

²⁰⁷ Hugh Clapin, “Introduction.” In *Philosophy of Mental Representation*, edited by Hugh Clapin (New York, Oxford: Oxford University Press, 2002) p. 1; Tim Crane, *The Mechanical Mind* (London, New York: Routledge, 2003), pp. 27-29.

²⁰⁸ Clapin, “Introduction,” p. 1.

²⁰⁹ Ibid., p. 1. See especially Hilary Putnam, *Reason, Truth and History* (Cambridge, New York: Cambridge University Press, 1981), pp. 1-2; Fred I. Dretske, *Knowledge and The Flow Of Information* (Stanford, CA: CSLI Publications, 1999; originally published: Cambridge, Mass.: MIT Press, c 1981); Jerry Fodor, *A Theory of Content and Other Essays* (Cambridge, Mass.: MIT Press, 1992); David Papineau, “Representation and Explanation,” *Philosophy of Science* 51, 4 (December, 1984), pp. 550-572; David Papineau, *Reality and Representation* (Oxford UK, Cambridge USA: Blackwell, 1987); David Papineau, *Philosophical Naturalism* (Oxford UK, Cambridge USA: Blackwell, 1993); Ruth Millikan, “Biosemantics,” *Journal of Philosophy* 86, no. 6 (June 1989), pp. 281-297; Ruth Millikan, *Language, Thought and Other Biological Categories: New foundations for realism* (Cambridge, Mass.; London, England: MIT Press, 1984); Robert Cummins, *Meaning and Mental Representation* (Cambridge, London: MIT Press, 1989).

Consider any basic thought like the desire to go to the school. In a very simple desire like this, one's mind may go through many relevant thoughts. I may think of the school bag in front of me, and I may believe that if I want to go to school, I should bring the bag to the school. I may consider whether I should put some notebooks for taking notes at the class next day. I may remember how enlightening and enjoyable was the previous class, especially remember the lecturer's interpretation of Gregor Samsa's change of mind/body as a reaction to the capitalist/totalitarian system. I may also calculate the travel expense from home to the school, how much it would cost me to go to school, whether I am prepared to accept the fact that on the condition of choosing to go to school, as a poor student, I will trade having a full stomach with having the opportunity for enlightenment. I can really think about many things, things that are present in my room now, things that exist somewhere but not in my room, things in the future, things happened in the past, and also things that have never existed and will never exist.²¹⁰ My mind is, most of the time, directed towards or about something in the world. In Brentano's terminology, this is called "intentionality", and it is, according to Brentano, what draws the sharp line between the mental and the material.²¹¹

²¹⁰ For a nice introduction to intentionality, see Hugh Clapin, "Introduction," pp. 1-2; Crane, *The Mechanical Mind*, pp. 30-41. For an introduction to its historical background, see Tim Crane, "Intentionality." In *Routledge Encyclopedia of Philosophy*, edited by Edward Craig (London, New York: Routledge, 1996), vol. 4.

²¹¹ Most people now disagree with Brentano's thesis. Nevertheless, they all say that intentionality is important and it should be given an account within a naturalist framework. For a discussion of Brentano's thesis in the context of contemporary philosophy of mind, see Crane, *The Mechanical Mind*, pp. 36-40.

One of the effective ways to describe intentionality is that intentional systems differ from non-intentional systems in that states of the former have the capacity to represent things in the world as being in a certain way, whereas states of the latter do not have any capacity to represent anything at all. At this point, the central issue of intentionality appears to be based on the very idea of representation: “[H]ow can a mere mechanism think about and represent things?”²¹²

The Idea of Representation

No one, including Ryle, probably rejects the idea that subjects represent the world. If I have a belief that it is snowing now, then I am in a certain state of mind that represents the world as being a certain way. It seems clear that it is simply non-sense to suppose that one could believe without believing anything. “What you believe or desire is what is represented by your belief and desire”.²¹³ Of course, this does not force one to accept a further claim that there exist mental representations in the mind, mental structures, which constitute intentional states and stand for states of affairs, objects and relations in the world.²¹⁴

According to Ryle’s view, this way of understanding the nature of mind creates insurmountable difficulties rather than providing an explanation of intelligent behavior. For it always brings forth the

²¹² Crane, *The Mechanical Mind*, p. 6.

²¹³ Crane, *The Mechanical Mind*, p. 23.

²¹⁴ So, it seems that we have two options here: (i) any given mental state, X, represents something Y (object, event, relation, situation, fact, etc.,) about the world. Ryle and Wittgenstein, as far as I understand, could presumably go easy with this option; (ii) any given mental state, X, represents Y in virtue of X’s constitutive elements, which are said to be “mental representations” in the mind.

questions like “what/who use those mental representations?” Ryle believes that nothing is intrinsically a representation of anything; so if we talk about mental representations, then we should also talk about the user or interpreter of mental representations. We must presuppose a homunculus, inner eye in the mind that uses and interprets those mental representations. This, in turn, leads to the postulation of another homunculus observing and making sense of the behavior of the former homunculus, and the process eventually leads to an infinite regress. This argument is sometimes called “Ryle’s infinite regress machine”, or interchangeably “explanatory regress”.²¹⁵

On the other hand, a group of philosophers argued for the claim that Ryle’s concerns are far from being warranted. With the technical and theoretical developments in computer and cognitive science, it became possible to give a computational/functional model of mind in which cognition/thinking is realized by mindless manipulators of internal representations of a computing machine.²¹⁶ Dennett, for instance, believes that it is possible to avoid being prey to Ryle’s infinite regress machine via a top-down research strategy in which each global intentional characterization of an intelligent system is analyzed in terms

²¹⁵ Gilbert Ryle, *The Concept of Mind* (London: Penguin, 2000), pp. 31-32; Daniel C. Dennett, “Styles of Mental Representation.” In *The Intentional Stance* (Cambridge, London: MIT Press, 1987), pp. 213-225; Daniel C. Dennett, “Artificial Intelligence as Philosophy and as Psychology.” In *Brainstorms: Philosophical essays on mind and psychology* (Cambridge, Mass.: MIT Press, 1978), pp. 119-122; Robert Moore, “A Cognitivist Reply to Behaviorism,” *The Behavioral and Brain Sciences* 7, no. 4 (December 1984), pp. 637-639.

²¹⁶ Many popular names were invented to portray the tendencies and the trends of this intellectual movement in the 1950s. Among the popular ones come “the cognitive turn,” and “the cognitive revolution.” For a nice introduction to the debate between behaviorists and cognitivists, see Robert A. Wilson, “Philosophy.” In *The MIT Encyclopedia of the Cognitive Sciences*, edited by Robert A. Wilson and Frank C. Keil (Cambridge, Mass.; London, England: The MIT Press, 1999), p. xxviii.

of its local sub-systems, and these sub-systems are then examined in terms of their own constitutive elements. At the end, Dennett believes, you may get to the level where nothing explicitly intelligent remains.²¹⁷ According to Dennett, we make use of these sorts of intentional characterizations in everyday life. For instance, about a chess-playing computer, a computer programmer can describe the machine by saying “it always moves the queen too early”. This might be taken to describe the computer as having intentional traits on the global scale, yet as you go down from the global scale to local scale you will not find anything explicitly intelligent once you come to grips with its details about sub-systems and implementation methods. So then, one can talk about inner (mental) representations without being a prey for Ryle’s infinite regress machine.

I am not really sure whether a computer model or top-down strategy can help solve the problem addressed by Ryle. From my point of view, there is a more basic, but very efficient way to deal with Ryle’s infinite regress machine. It seems that Ryle begins with the assumption that nothing is intrinsically a representation of anything else. This does not appear to me as obvious as it does for Ryle. One may plausibly say that there are different kinds of representations. In conventional representations, for instance, the assumption seems to have a *prima facie* validity. For instance, a red traffic light, for a pedestrian, represents “stop”. But, this is simply a matter of convention. On its own merit, a red traffic light does not mean or represent anything. It is only meaningful within the cultural and social conventions of a society. So, it represents

²¹⁷ Dennett, “Artificial Intelligence as Philosophy and as Psychology,” p. 124.

“stop” only for an interpreter/user who is bound to those values and understands the normativity behind rules of a society.²¹⁸ However, mental representations seem to be different from conventional representations such that no interpretive activity is actually required. For instance, consider the behavior of a new born human being (the behavior of calling for attention each time she sees her mother). Upon seeing and smelling her mother for the first time, and thereby forming a mental representation of her in terms of her visual and olfactory qualities, are we supposed to say that this is no different kind of representation from that of conventional types? It may really be very wrong to believe, as a general claim, that each kind of representation requires some interpreter in order to be a representation of something. So, denial of Ryle’s initial assumption can be a more effective way to deal with Ryle’s infinite regress machine.²¹⁹

This (direct) denial also comes with a reward, something that has been recurrently emphasized and appealed to in the history of philosophy probably since Aristotle. Take one of the basic, but most essential characteristics of having a mind, for instance, the navigational capacity. I am a living mental being, and part of being a living mental

²¹⁸ Crane, *The Mechanical Mind*, pp. 175-176.

²¹⁹ This line of thinking can be found in Frank Jackson, “Representation and Experience.” In *Representation in Mind: New Approaches to Mental Representation*, edited by Hugh Clapin, Phillip Staines, Peter Slezak (Amsterdam; Boston: Elsevier, 2004), pp. 109-110. Jackson points out the contrast between the way conventional representations (maps, diagrams, sentences and so forth) represent and the way perceptual experience represents. The crucial difference, according to Jackson, is that while the former always requires some external agent for its use and interpretation, the latter only requires that one has a perceptual experience of something. Nothing else is required, contrary to Ryle’s worries. So, there is a difference in kind between mental representations and conventional representations. This might be one way of dealing with Ryle’s infinite regress machine.

being is to store mental representations about the environment and use them to navigate and fulfill my needs on this earth, most importantly the inevitable need to find food and not die out of hunger. Again, one can take the phenomenon about the migration of birds as one of the most striking instances for a navigational capacity. It is reported that some birds “represent the configurations of constellations in the night sky, using these celestial maps to find their way on their long flights”²²⁰ C. R. Gallistel makes the following interesting observation:

On the featureless Tunisian desert, a long-legged, fast-moving ant leaves the protection of the humid nest on a foraging expedition. It moves across the desert in tortuous loops, running first this way, then that, but gradually progressing ever farther away from the life-sustaining humidity of the nest. Finally it finds the carcass of a scorpion, uses its strong pincers to gouge out a chunk nearly its own size, then turns to orient within one or two degrees of the straight line between itself and the nest entrance, a 1-millimetre-wide hole, 40 metres distant. It runs a straight line for 43 metres, holding its course by maintaining its angle to the sun. Three metres past the point at which it should have located the entrance, the ant abruptly breaks into a search pattern by which it eventually locates it. A witness to this homeward journey finds it hard to resist the inference that the ant on its search for food possessed at each moment *a representation of its position* relative to the entrance of the nest, a spatial representation that enabled it to compute the solar angle and the distance of the homeward journey from wherever it happened to encounter food.²²¹

²²⁰ Michael Strevens, “The Explanatory Role of the Notion of Representation,” p. 6. Available online: <http://www.strevens.org/research/expln/mimeoexpln.pdf>. In a similar vein, one can also think of many similar cases of (apparently) more intelligent beings like human beings. Consider, for instance, the behavior of a sailor in the old times who does not have any compass. Knowing that the sun rises in the east and sets in the west could have helped him to determine the direction. Or, when it is night, he could simply locate the North Star in the sky, and by following it, he could find the way back to wherever he needed to go.

²²¹ C.R. Gallistel, *The Organisation of Learning* (Cambridge, Mass.: MIT Press, 1960), p. 1. Italics belong to me.

The most important thing about this observation and the experiments following the observation is that the ant has a mental representation of its position in relation to the direction to the nest, and this mental representation has been constantly updated as it moves around the desert.²²² The idea of postulating mental representations has, therefore, an important explanatory role in describing and predicting the behavior of mental beings. Mental representations are not just sums of structures in the head purported to describe how things stand in the world; they also have the causal-explanatory import for the interaction between mind and world. Strevens nicely puts the whole explanatory value of mental representations in a Kantian dictum: “Action without accurate representation is blind; accurate representation without appropriate action is useless”²²³

I am a believer in this Kantian dictum. The philosophical question, however, is to understand and spell out the very nature of those mental representations: What does determine the ground of mental representations? In virtue of what can an inner state of a subject (of mind) represent and think about the world? More generally, what is it to have a mind? What is it to have mental states?

From Functionalism to Post-Functionalism

There have been some influential attempts to answer this question in the twentieth century philosophy of mind. Let us go back to the previous

²²² Strevens, “The Explanatory Role of the Notion of Representation,” p. 5; Crane, *The Mechanical Mind*, pp. 80-81.

²²³ Strevens, “The Explanatory Role of the Notion of Representation,” p. 7.

chapter and remember what functionalism exactly says. Functionalism rejects type-identifying mental states with neural states and denies the non-causal, anti-realist account of mind given by behaviorism, and eventually it offers a very liberal account of mind according to which mind is identified as a functional kind (of appropriate complexity).

Putnam, the father of this ideology, was very much affected by technical and theoretical breakthroughs in computer science at that time, and he proposed a very interesting and different account of mind. For this account, all psychological states must be understood and explained in terms of computational / functional states.²²⁴ So, for instance, “a human being is just a computer that happens to be made of flesh and blood, and that the mental states of a human being are its computational states”.²²⁵

More precisely:

Any being / system B is in a type M of mental state, iff there is a state of B that plays a unique functional / causal role F definitive of M-type within a complex network of states of B such that it mediates, together with other relevant states in that network of states, between perceptual inputs and behavioral outputs of B.

As I said, this was supposed to explain all psychological states (e.g., pains, beliefs, desires, hopes, and so forth). Thus construed, functionalism does not only say what kind of thing mind is; it also tells something

²²⁴ In his article “Psychological Predicates”, Putnam only gives the example of pain, but his real purpose was to show that if this functionalist strategy can be applied to pain, it can equally be applied to the rest of psychological states such as belief, desire, fear and so forth. See Hilary Putnam, “Psychological Predicates.” In *Art, Mind, and Religion*, edited by W. H. Capitan & D. D. Merrill (University of Pittsburgh Press, 1967), reprinted later with the new title “The Nature of Mental States” in *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980a), pp. 223-231. For the same point, see also Hilary Putnam, “Functionalism: Cognitive Science or Science Fiction?” In *The Future of Cognitive Revolution*, edited by David Martel Johnson and Christina E. Erneling (New York, Oxford: Oxford University Press, 1997), p. 33; Hilary Putnam, “Putnam, Hilary.” In *A Companion to the Philosophy of Mind*, edited by Samuel Guttenplan (Cambridge, Mass.: Blackwell, 1994), pp. 507-513.

²²⁵ Putnam, “Functionalism: Cognitive Science or Science Fiction?” p. 32.

about the individuation of representational/intentional states. For me to represent, for instance, the fact that the school bag is in this room, I should have been affected by the presence of a school bag in the room, and this should have then caused a state of mind, whose function, together with the other relevant states in the network, is to elicit a certain kind of behavior relevant to that stimulus/input. For me to represent the fact that I am in pain, I must have received bodily/spiritual damage (e.g., being hit by a material being, being disturbed by a spiritual being) to my functional organization/network, and this should have led to certain inner states (e.g., the belief that I am in pain, the desire to get rid of pain, and so forth) together which elicit a certain kind of output (e.g., the sound “ouch”, spiritually getting away from the source of psychological disturbance, and so forth).

Putnam particularly emphasized that his theory of mind is so liberal with respect to the physical make-up of mental beings that even, for instance, “We could be made of Swiss cheese and it wouldn’t matter”.²²⁶ So, it is a purely empty and futile enterprise, according to the functionalist metaphysics of mind, to explore the source of psychological laws, which govern mental operations and behaviors of material or spiritual beings, in the underlying (material/spiritual) structure/substance where those laws make their first appearance. This is another way of saying that there is no difficulty, in principle, for the individuation of the same mental states in physically/spiritually differently composed beings and substances, provided that those beings

²²⁶ Hilary Putnam, “Philosophy and Our Mental Life.” In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block (Cambridge, Mass.: Harvard University Press, 1980), p. 134.

and substances manifest functionally equivalent architectures, i.e., a functional organization of an appropriate kind that is constitutive of mentality and intelligence. Basically, what matters is the functional organization between inner states, sensory inputs and motor outputs. So, according to the former self of Putnam, had there been creatures made of Swiss cheese that possessed functional organization of appropriate complexity, it would have been able to think and represent the cheesy world wherein it is situated. Let this suffice as a quick tour back to functionalism.

As the years passed by, the latter self of Putnam gradually began to recognize that his former self was fundamentally wrong about the way to explain intentional/representational states via a functionalist account, one that he proposed as the best theory of mind back then in the late 60s.²²⁷ This change of mind, for Putnam, has occurred in his notorious article "The Meaning of 'Meaning'".²²⁸ In reporting the reasons for his recantation of functionalism, he says:

Ever since I wrote "The Meaning of 'Meaning'," I have defended the view that the content of our words depends not just on the state of our brains (be that characterized functionally or neurophysiologically), but on our relations to the world, on the way we are embedded in a culture and in a

²²⁷ See Putnam, "Functionalism: Cognitive Science or Science Fiction?" pp. 32-44. In the opening paragraph of this essay (p. 32), Putnam says: "There is an ancient form of 'functionalism'—so Martha Nussbaum and I have argued—that can be found in Aristotle's *De Anima*. This is the view that our psyches can best be viewed not as material or immaterial organs or things but as capacities and functions and ways we are organized to function. In that wide sense of the term, I am still a functionalist. In this chapter, however, I will be considering a contemporary rather than an ancient way of specifying what it is to be a functionalist, one I myself introduced in a series of papers beginning in 1960."

²²⁸ Hilary Putnam, "The Meaning of 'Meaning.'" In *Language, Mind and Knowledge*, edited by K. Gunderson. Minnesota Studies in the Philosophy of Science, Vol. 7. (Minneapolis: University of Minnesota Press, 1975), pp. 131–193. Reprinted in Hilary Putnam, *Mind, Language and Reality, Philosophical Papers, Volume 2*. (Cambridge: Cambridge University Press, 1975), pp. 215–271.

physical environment. A creature with no culture and no physical environment that it could detect outside its own brain would be a creature that could not think or refer, or at least (to avoid the notorious issue of the possibility of private language) could not think about or refer to anything outside itself.²²⁹

This article was a kind of turn in Putnam's understanding of mind. Let's remember what it was all about. Suppose that there are two worlds, Earth and Twin-Earth. For everything in Earth, we have the corresponding molecular-to-molecular duplicate of the same thing in Twin-Earth. So, suppose that, for instance, Oscar lives in Earth. Then, there will be a Twin-Oscar (shortly as T-Oscar) in Twin-Earth. Suppose also that everything is identical between these two worlds except the molecular structure of "water". So, we have H₂O in Earth, and XYZ in Twin-Earth. Putnam argues for the claim that Oscar and T-Oscar will have different psychological states about "water", because their environments are different. When Oscar utters the word "water", what he means by / refers to will be H₂O, whereas when T-Oscar utters the word "water", what he means / refers to will be XYZ. So, their water-thoughts will be different. Does this argument show that functionalism is incorrect? For Putnam, it does show, in a way, that we are getting away from the ideology of original functionalism:

Mere computational relations between speech events and brain events do not, in and of themselves, bestow any content whatsoever on a word, any more than chemical and physical relations do. But this implies that no mental state that has content (no "propositional attitude") can possibly be identical

²²⁹ Putnam, "Functionalism: Cognitive Science or Science Fiction?" p. 36.

with a state of the brain, even with a computationally characterized state of the brain.²³⁰

The problem, as far as I understand, seems to be the following one: The account of meaning/content given in “The Meaning of ‘Meaning’” seems to be in conflict with the account of content given in “Psychological Predicates”. But, the situation does not seem to be hopeless, at least for now. Putnam says that we can still have the option to ignore all external factors (e.g., culture, environment, and so forth) and look for an abstraction, something that explains what makes any type of psychological state the kind of state it is on a very narrow level without appealing to external factors. So, as opposed to “wide content”, we may now have a new notion called “narrow content”. Putnam says:

The advantage of the suggestion, from a functionalist point of view, is that "narrow content," by definition, has to do only with factors inside the organism; thus, there is at least the hope that narrow contents might be identifiable with computational states of the organism, thus realizing a version of the original functionalist programs. But the suggestion has problems.²³¹

The notion of “narrow content” is quite a problematic notion, and it is still a controversial issue whether anyone has given an independent and clear definition of “narrow” content, one that is not parasitic on the wide content and one that all psychological states have in common.²³² Without having any way to make sense of “narrow content”, then one could accept the fact that “narrow content” is problematic and decide to stick

²³⁰ Ibid., p. 36.

²³¹ Ibid., p. 36.

²³² Ibid., p. 37.

only with “wide content”. So, under this alternative approach, we can state functionalism as a theory of mind that explains representational/intentional states via reference to the community and the environment. So, now we move from a narrow functionalism, which was the original theory, to a wide functionalism. According to wide functionalism, then “mental states are identical with computational-cum-physical states of organisms plus communities plus environments”.²³³ For Putnam, this means nothing but departing away from the original functionalist hypothesis, a move that renders wide functionalism simply useless:

Of course, one might decide to drop the notion of "narrow content," and say: "Very well then. If mental states are individuated by contents which are themselves partly determined by the community and the environment, then let us widen the functionalist program, and postulate that mental states are identical with computational-cum-physical states of organisms plus communities plus environments. But how useful is it to speak of "computational-cum-physical states" of such vast systems?"²³⁴

I have charged functionalism with utopianism, with being "science fiction" rather than the serious empirical hypothesis that, in my early papers, I hoped to provide. I have also suggested that bringing in such concepts as "narrow content" and "conceptual role" only drains the functionalist proposal of its original substance—it turns it into a case of using concepts that stand for we-know-not-what as if they had serious scientific content.²³⁵

As far as I understand, with his article “The Meaning of ‘Meaning’”, the latter self of Putnam recognized the gravity of a destructive dilemma

²³³ Ibid., p. 37.

²³⁴ Ibid., p. 37.

²³⁵ Ibid., p. 38.

troubling his former self. The dilemma is the following: (1)

Functionalism, at bottom, is a theory of mind that is purported to explain mental states without appealing to different parameters surrounding the subject such as environment, culture, politics, etc. For instance, the virtue of original functionalism is to explain the mental state of being in pain without any particular reference to the environment, physical/spiritual make-up and so forth (remember my example about spiritual and material pain). This presupposes that there is an ideal psychological theory that can give a unified account for all pain-capable feeling beings/systems. So, there must be a “narrow” content that makes something the kind of mental state it is. But, the very notion of “narrow content” is very problematic. No one knows what “narrow content” is exactly; hence Putnam says “it turns it into a case of using concepts that stand for we-know-not-what as if they had serious scientific content.” (2)

On the other hand, we may still have the option to appeal to “wide content”. However, now we must add many different parameters in our basic functionalist account to describe any given representational/intentional state. This, according to Putnam, leads us far away from the original functionalism. In particular, Putnam does not see any use in taking such an approach. At the end, it seems for later Putnam that original functionalism was completely a utopian enterprise, and it should be abandoned, if the target is to create something scientific, not science-fictional. Of course, not all people agree with Putnam. After “The Meaning of ‘Meaning’”, post-functionalist philosophers began to appear on the stage:

Of course, those who are sympathetic to functionalism have not given up as a result of my recantation; there are a number of what we might call "post-functionalist" programs on the market. One kind of post-functionalist program seeks to avoid the difficulties inherent in the "implicit definition by a theory" idea that was at the heart of classic functionalism, by relying entirely on external factors to fix the contents of thoughts.²³⁶

As I have said in the beginning of this chapter, the big question between the 1980s and 1990s was the mystery of intentionality. Putnam, with his article "The Meaning of 'Meaning'" only showed, or convinced many people, that there is not much hope in narrow functionalism. Post-functionalist philosophers then began to work on a completely externalist, wide account of functionalism, one that explains intentional/representational states through reference to the environment. In this thesis, I gave a space to the views of two philosophers from the post-functionalist programme. Dretske and Millikan's accounts in the following chapters should not be read only as different theories of mental representation. Their theories are also meant to promote the idea that representation can still be understood within the largely functionalist approach.

In this chapter, I began with a brief historical tour into the philosophical climate of the twentieth century philosophy of mind. Intentionality was one of the central notions in the literature, and it seems, according to many people, that intentionality boils down to the very notion of representation. I talked about two different ways of understanding representation. Then, I revisit functionalism one more time, because it tells us something about the individuation of intentional/representational states. In order to understand the very idea

²³⁶ Ibid., p. 39.

behind post-functionalist philosophers such as Dretske and Millikan, it was vitally important to make some room for understanding Putnam's change of mind. That's why I explained at some length one of the reasons that led later Putnam to abandon the original functionalist hypothesis and led others, like Dretske and Millikan, to renovate functionalism with a new energy.

CHAPTER V

DRETSKE'S ACCOUNT OF REPRESENTATION

In this chapter, I will deal with Dretske's account of information and representation. In the first section, I will provide a brief presentation of Dretske's account as it stands in his most influential book *Knowledge and The Flow Of Information*.²³⁷ After laying out the basics of Dretske's account, I will present, in the second section, two chief troubles with his account and how he deals with these issues. In the end, I will explain the reasons for the failure in his solutions.

Dretske's Theory

Fred Dretske, one of the important figures in the literature on theories of content/representation, proposes a new perspective that satisfies, according to his point of view, the requirements of naturalism in both epistemology and the philosophy of mind by supplying a powerful resource that can explain notions such as knowledge, belief and experience within the vocabulary of natural sciences. Not only it does bring forth a new perspective, it also presents a new way of dealing with the problem of misrepresentation in the literature.

Let us begin with the notion of information and attempt to grasp what information is and is not from Dretske's point of view. The received common view, Dretske says, argues for the claim that "something only

²³⁷ Fred I. Dretske, *Knowledge and The Flow Of Information* (Stanford, CA: CSLI Publications, 1999; originally published: Cambridge, Mass.: MIT Press, c 1981). Henceforward shortly as KFI.

becomes information when it is assigned a significance, interpreted as a sign, by some cognitive agent".²³⁸ Accordingly then, the very idea of information as something independent of interpreters was taken to be an absurd idea. Information must be in the head of the receiver much like the beauty is in the eye of beholder. Thus, from this approach, there is no such thing as a flow of information between two events if there remains no intelligent life on earth to observe and interpret the relation of those two events as a sign of information.²³⁹

From Dretske's point of view, however, the received view is based on a common confusion of the idea of information with the idea of non-natural meaning. Information is a commodity that is quantifiable and objective, whereas non-natural meaning is not something that one can quantify over and it definitely requires subjects. Paul Grice, in his very famous and often referred to article "Meaning",²⁴⁰ makes a distinction between natural meaning and non-natural meaning. According to him, non-natural meaning has to do with the language and semantics where a speaker's intentions are important to fix the content of his utterance. If a subject utters a sentence "I have the measles", then, according to Grice, he typically intends to use this utterance in order to communicate what he thinks he has to the audience.²⁴¹ Of course, one may mean that he has the

²³⁸ Ibid., preface vii.

²³⁹ Cf. Daniel C. Dennett. "Artificial Intelligence as Philosophy and Psychology." In *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge, Mass.: MIT Press, 1981), pp. 109-126. In this article, Dennett deals with a very similar question whether representation is something that exists only when there is an intelligent being that can interpret and use that representation.

²⁴⁰ Paul Grice, "Meaning," *Philosophical Review* 66, no. 3 (Jul., 1957), pp. 377-388.

²⁴¹ Ibid., p. 383.

measles but not really have them at all. In contrast, natural meaning is independent of language and intentions of a speaker. Think of the causal law-like relations between two types of events. Events of type A always cause events of type B. Whenever there are spots on somebody's face it is always because of the measles. Here, "those spots mean measles" is an instance of natural meaning. Moreover, if x naturally means y, then this indicates that there is a causal law-like relation between x and y, then y and x must be present. In other words, if spots mean_{naturally} measles, then "spots mean_{naturally} measles" entails the presence of measles and spots.

In this context, Dretske believes that information must be distinguished from non-natural meaning in Grice's sense, and it must be associated with natural meaning. X's fingerprints on a glass of water at the crime scene naturally mean that X was present at the crime scene. They indicate or carry information about the presence of X at the crime scene. The number of rings on the trunk of a tree naturally means the age of the tree. It indicates or carries information about the age of the tree. This is the ordinary sense, or nuclear sense, of the term "information" from Dretske's point of view.²⁴² Understood in this manner, false information and misinformation, or synonymously as false indication or misindication,²⁴³ are not kinds of information. For if there is an informational relation between two types of events, say the dance of a honeybee and the whereabouts of the nectar, then this entails the truth about the existence of the nectar. If a scientist tells us a very important fact about the relation between the number of rings in the trunk of a tree

²⁴² Dretske, *Knowledge and The Flow Of Information*, p. 45.

²⁴³ Dretske uses the term "indication" as the synonym for the term "information".

and the age of that tree, he definitely makes use of the nuclear sense of information where the informational relation between the number of rings and the age of the tree entails the existence of the tree. The essential connotation of the term “information”, or “indication”, is related to the truth.²⁴⁴ Therefore, misinformation and false information, in Dretske’s account, are not kinds of information. Moreover, it will be redundant to speak of information as being reliable because misinformation and false information are already excluded from the extension of the concept “information”.²⁴⁵

“In the beginning there was information. The word came later,” says Dretske.²⁴⁶ These are the first two sentences of his book KFI, and they actually give important hints about the big project in his mind. From Dretske’s point of view, information is an “objective commodity, something whose generation, transmission, and reception do not require or in any way presuppose interpretive processes”.²⁴⁷ The project in KFI, in this context, is to explain the transition from information to meaning that “was achieved by the developments of organisms with the capacity for

²⁴⁴ Dretske, *Knowledge and The Flow Of Information*, pp. 45-6. In everyday life, we actually take misinformation and false information as a species of information. For instance, we might have a very extensive piece of information about the location of the Blue Mosque, but it may turn out later that that piece of information leads us somewhere other than the Blue Mosque. So, we say that it was a very good piece of information except for the fact that it mistakes Saint Sophia for the Blue Mosque. It was a wrong piece of information. Dretske, however, does not want to totally reduce information to its everyday usages, because he believes that the essence or the function of information must be the one that should provide something capable of yielding truth and knowledge. From his point of view, if information were not taken to be related to the truth, then this would automatically make all the hard work and troubles to retrieve information from relevant sources futile and meaningless. Intelligence services of a country, for instance, would not be needed.

²⁴⁵ Dretske, *Knowledge and The Flow Of Information*, p. 45.

²⁴⁶ Ibid., preface vii.

²⁴⁷ Ibid., preface vii.

selectively exploiting this information in order to survive and perpetuate their kind".²⁴⁸ So, for Dretske, the central task in KFI is to explain how meaning can evolve, how higher-order cognitive mechanisms can be seen as the development of lower-order information-processing mechanisms.²⁴⁹ Dretske says:

The higher-level accomplishments associated with intelligent life can then be seen as manifestations of progressively more efficient ways of handling and coding information. Meaning, and the constellation of mental attitudes that exhibit it, are manufactured products. The raw material is information.²⁵⁰

Suffice it to say, then, the initial aim in KFI is to explain how meaning can be considered as the evolutionary product out of the raw material as information. Of course, this should not be thought as the easy task, because it does seem that there is an insurmountable gap between information and non-natural meaning. Information, as understood in the mathematical sense, is objective and quantifiable, whereas non-natural meaning is always understood within the context of a subject's mind-world relationships and the quantitative expression "amount of meaning" simply does not make sense.²⁵¹ The Communication Theory

²⁴⁸ Ibid., preface vii.

²⁴⁹ Ibid., preface vii; Fred I. Dretske, "Précis of Knowledge and the Flow of Information," *Behavioral and Brain Sciences* 6, no. 1 (1983), pp. 55-63. By using the word "evolution", I want to emphasize that the difference between simple information-processing mechanisms such as TV sets, radios, etc., and complex information-bearing beings such as vervet monkeys, human beings, etc., is only a matter of degree from Dretske's point of view—this point will become more clear in the following pages. In this respect, Dretske and Dennett share the same view that, contrary to Brentano's view, there is not any principled difference between mental beings and physical beings. Cf. Daniel C. Dennett. "Intentional Systems in Cognitive Ethology: The "Panglossian Paradigm" Defended." In *Intentional Stance* (Cambridge, Mass.: MIT Press, c1987), pp. 237-268.

²⁵⁰ Dretske, *Knowledge and The Flow Of Information*, preface vii.

²⁵¹ Ibid., pp. 42-3.

(henceforward shortly as CT), or interchangeably called “The Mathematical Theory of Information”, for instance, deals exclusively with the determination of the amount/quantity of information generated at the source of information and how much of this information is transmitted to the receiver.²⁵² For engineering purposes, calculation of the statistical averages about the flow of information between two points can be one of the important tasks, and CT helps engineers to design and create systems with effective capacities for handling and delivering information. Of course, here the important thing is about the amount of information, not its content. As for non-natural meaning, it simply does not make sense to investigate things like how much meaning is produced in two statements, how much meaning is carried between two people. We are simply interested in the content/meaning of statements, not sorts of things CT is interested in. So, given these difficulties, it appears that CT alone cannot explain the transition from informational measure to informational content/meaning.

CT is at the center of Dretske’s information-theoretic account, so I would like to present it briefly without entering into technical details. Let us proceed with Dretske’s example. Our task is to calculate how much information is generated and transmitted. The boss pays our office a visit and utters his decree “one of the employees out of eight here at this office will be selected for a secret mission and I will leave the decision procedure to you”. It is equally likely for each one of the eight employees to be selected for this secret mission. This group of employees can make

²⁵² Ibid., p. 3.

use of quite different decision procedures, such as flipping a coin, drawing straws, partitioning the group into various natural classes, to pick up one individual in the group. But, CT will always compute three bits of information at the end of each decision procedure because of the formula $I_{(s)} = \log n$.²⁵³ In other words, CT will not look at who is selected, but at how much information is produced in each case and why there is always the same amount of information at the end of each decision method.

[S]ince the mathematical theory of information deals exclusively with the quantitative aspects of these situations, exclusively with amounts of information, the result, from the point of view of this theory, would be the same whether Margaret or Herman was the nominee. In both cases the result of the selection procedure, and the note to the boss, would contain exactly the same amount of information. This theory does not provide a way of distinguishing between what we intuitively take to be the difference between the information that Herman was selected and the information that Margaret was selected. This latter is a difference in the informational content of a state of affairs (or message), a difference that the quantitative theory ignores.²⁵⁴

CT is purely a quantitative theory meant to determine the amount of information generated at the source and how much of this information is received at other points.²⁵⁵ It cannot tell you what information is encoded in a signal, but only provides how much information it carries. In the

²⁵³ Here, "I" stands for the information, "s" for the source and "n" for the number of possibilities to 1. Granted that the base of logarithm is two, when we apply this general formula for computing the amount of information to the case in point, $I_{(s)} = \log 8$, we will get the result that three bits of information is generated at the source. For instance, let us say that the amount of information is always coded in binary digits (0,1). At the end of the decision procedure above, we will have exactly three bits of information (e.g., 101). In my presentation, I will skip the technical niceties of CT in handling the situations where not all cases are equally likely.

²⁵⁴ Dretske, *Knowledge and The Flow Of Information*, p. 6.

²⁵⁵ Ibid., p. 3.

light of this observation, it is solidly evident that CT is very limited to help Dretske explain the evolution of meaning from the information-processing mechanisms and make good sense of Dretske's big naturalist project that mind is the biggest consumer of information.²⁵⁶ It is conceivable, as far as I observe Dretske's big project, that a being/system might have had much more fruitful and richer resources to record various details of visual information compared to human beings' capacity to store visual experience and yet not able to manifest intelligence such as experience, belief and knowledge and thereby not go through a process of intellectual evolution at all.²⁵⁷ In other words, it is not important for a being/system to have a capacity of great storage in order to instantiate intentional properties such as belief, experience and knowledge.²⁵⁸ The important point is to explain how a relatively low-order information-processing mechanism can evolve into a more sophisticated and intelligent being. In order to explain this transition, I guess this is what Dretske has in mind, one needs to explain the informational content, not the informational measure, of a signal or state of affairs, whatever have you, and then move on to discussion of how this raw, messy and indeterminate informational content contributes to the evolution of

²⁵⁶ Fred I. Dretske, *Perception, Knowledge and Belief* (Cambridge, U.K.; New York: Cambridge University Press, 2000), preface xi-xii.

²⁵⁷ I take that one of the targets of Dretske's distinction between analog and digital information is to underline, albeit very implicitly, this conceivability argument and explain the real reason for the evolution of information-processing mechanisms towards more sophisticated intelligent beings. I will come back to this point later on in this chapter. I designed this conceivability argument after my readings of Dretske's book *Knowledge and The Flow Of Information*, and this is actually my interpretation of what he aims to do in this book. As far as I know, he does not make explicit use of this argument anywhere in the book, but I believe it is hidden between the lines.

²⁵⁸ Dretske, *Knowledge and The Flow Of Information*, preface vii.

intelligent beings. Dretske believes that if CT is supplemented with some further conditions and ideas, it can become a genuinely semantic theory that can help in semantic and cognitive research.²⁵⁹ So, in a nutshell, Dretske's book KFI is an attempt to develop a philosophical theory of information explaining the semantic content of mental states without invoking intentional terms.²⁶⁰

After having laid down the principles and features of his theory, especially about the decisive line of demarcation between informational relationship and causal relationship on the one hand and the semantic conditions upon which a philosophically interesting theory of information is securely built on the other,²⁶¹ Dretske finally presents a theoretical definition of informational content:

"Informational content: A signal r carries the information that s is F = the conditional probability of s 's being F , given r (and k), is 1 (but, given k alone, less than 1)."²⁶²

²⁵⁹ Ibid., preface x. Later on, Dretske found out that his view "information-based naturalization of knowledge and perception" was not enough. He believes that you cannot expect to naturalize epistemology without naturalizing the mind and we need to explain what belief and experience are in more details than the one briefly offered within the project in *Knowledge and The Flow Of Information*. For this line of thinking, see Dretske, *Perception, Knowledge and Belief*, preface ix-x.

²⁶⁰ I owe great thanks to Fred I. Dretske for his comments on the earlier version of this paragraph.

²⁶¹ Dretske, *Knowledge and The Flow Of Information*, pp. 26-39; pp. 63-82. In the earlier drafts of this section, I was planning to make a room for the discussion of the difference between causal relationships and informational relationships as well as the discussion about the conditions for making a semantic theory out of CT. But then I realized that the whole enterprise is not practically feasible and I finally decided to leave that task out of the boundaries of this section.

²⁶² Ibid., p. 65. Take " k " as the background knowledge relative to the possibilities at the source. See also Dretske, "Précis of Knowledge and the Flow of Information," p. 58. Dretske's example: if I am playing chess and I receive the information that your knight is not on KB-3 (by some signal), then this signal could indicate to me that it's on KB-5 only if I already know that all the other possible positions that your knight could occupy are already occupied by other pieces.

Let us now go over some of the important features of informational content described by the definition above. First of all, it nicely captures the very idea behind CT. Any genuine communication supervenes on relevant objective relations that hold between two events. That is to say, if two events are lawfully related to each other, i.e., if there is a nomic regularity such that F-event is always followed by G-event, then it is clear that one can read off the information about F from the presence of G. G carries information about, or simply indicates that F is the case. Then, fingerprints of a person carry information about the presence of that person at the crime scene and can make him a strong suspect (given the background knowledge about his possible motivation to commit a crime in that scene, etc.); traces in the snow carry information about the species of living beings thereabouts; the dance of a honey bee indicates the presence of nectar-rich locations and sources.

Furthermore, the definition captures the ordinary sense of information that can be found in dictionaries where information is explained as something primarily and strictly related to intelligence, news, instruction and knowledge, each of which has been associated with the notion of truth.²⁶³ For Dretske, information is such “an important epistemic commodity” that people “buy it, sell it, torture people to get it, and erect booths to dispense it”.²⁶⁴ That’s why the conditional probability of s’s being F is equal to 1. In other words, no signal, event, or structure

²⁶³ Dretske, “*Précis of Knowledge and the Flow of Information*,” p. 57.

²⁶⁴ *Ibid.*, p. 57.

“can carry the information that *s* is *F* unless, in fact, *s* is *F*’.”²⁶⁵ In this framework, false information, misinformation and disinformation are not species of information just as a decoy duck is not a species of duck.²⁶⁶

In addition, the definition of informational content also leads us to the corollary, what Dretske calls the “Xerox principle”, according to which a signal’s informational content cannot be ever unique.²⁶⁷ Let us say that we have a signal transmitting a piece of information that *s* is *F* (say *F*=being a sphere) about a certain source of information. Let us also assume that there are objects at the relevant domain, i.e., at that source, which happen to instantiate *G*-properties (say *G*=being plastic). The Xerox Principle tells us that if there is a nomic regularity between *F*’s and *G*’s such that all *F*’s are *G*’s in that domain of objects, then any signal carrying the information that *s* is *F* necessarily carries the information that *s* is *G* as well.²⁶⁸ If there isn’t a nomic regularity, for instance if it is a weird coincidence that all *F*’s are also *G*’s, then a signal will carry the information that *s* is *F* without carrying the information that *s* is *G*, or vice versa.

Above all, the definition gives us a clue to figure out intentional characteristics of physical beings ranging from the ones whose information-processing units are relatively primitive to the ones whose information-processing capacities are very sophisticated and resourceful.

²⁶⁵ Ibid., p. 57.

²⁶⁶ Ibid., p. 57.

²⁶⁷ Ibid., pp. 57-8.

²⁶⁸ Dretske generally makes use of the following vivid example: the acoustic signal of a doorbell informs me that there is someone at the door, that the doorbell is being pressed, etc., provided that there is no short circuit between the doorbell and the doorbell button.

For instance, let us say my belief that this man is Susan's husband is different from my belief that this man is my cousin. Unbeknown to me, Susan's husband could turn out to be my cousin, so it would be wrong to substitute the content of my belief "this man is Susan's husband" with the co-referentially equivalent expression "this man is my cousin" *salva veritate*. In the same manner, Dretske conceives, the information that this man is Susan's husband can be taken to be different from the information that this man is my cousin. It simply gives us different pieces of information about the same source. In this regard, unavailability of the substitution of one piece of informational content with a co-referentially equivalent piece of information could be the outcome of different channels of communication carrying different pieces of information about the same source (plus the lack of nomic regularity between those two pieces of information). This is such a big challenge to Brentano, because Dretske makes it clear that there is no principled distinction between mental beings and physical beings. To put it in different words, Dretske does not believe that we are different from voltmeters, thermometers, etc., with respect to this feature of intentionality (unavailability of substitution); because they occupy intentional states very much like human beings. According to him, we are only different from them with respect to our degree of intentionality, we are different from them in the way we store, process, encode and utilize the same piece of information.²⁶⁹

²⁶⁹ Dretske, "Précis of Knowledge and the Flow of Information," p. 58.

Two Chief Troubles

Everything so far gives a hope about the bright future for informational account of content. Nonetheless, there are some big issues in Dretske's account that need to be resolved. First of all, it is typical of a belief that it does occasionally misrepresent how things stand in the external world. Our belief-structures are capable of misrepresenting the world around us, and this is one of the important features of our intentionality. Yet, according to the informational account, no structure can carry the information that *s* is *F* unless *s* is *F*; so there is no room for misrepresentation in this account. Misrepresentation is one of the salient features of intentionality, and the informational account seems to fail in this respect.²⁷⁰

Secondly, it turns out that the informational account of content misses the whole point about the failure of substitution of co-referentially equivalent contents and the real reason behind the conversion of truth-value. Substitution of one piece of information, according to Dretske's account, with another piece of information totally depends on a lawful relation between two events.²⁷¹ Turning back to the previous example, if

²⁷⁰ I haven't talked much about one of the big projects in KFI in details. In KFI, Dretske offers a new way of understanding mind and knowledge. For instance, instead of the traditional account of knowledge as "justified true belief", Dretske proposes a new account according to which knowledge is described as "information-produced belief" (See Dretske, "Précis of Knowledge and the Flow of Information," p. 55). In this whole framework, there are certain assumptions that play a key role in understanding the project in KFI. For instance, Dretske takes a person's beliefs as a certain kind of mental representations. Furthermore, Dretske's theory of informational content is supposed to explain the content expressed by belief-representations. These two points must be kept in mind, especially during his discussion of the problem of misrepresentation.

²⁷¹ Dretske, "Précis of Knowledge and the Flow of Information," p. 62. In Dretske's own words, "No structure can have the (informational) content that *s* is *F* without having the (informational) content that *s* is *G*, if it turns out that nothing can be *F* without being *G*."

spherical objects are all plastic objects in a given domain, and we receive a signal that an object in that domain is spherical, then this means we also receive the information that the object is plastic. If there is no lawful regulation between spherical and plastic objects, then substitution of one piece of information with a co-referentially equivalent piece of information will naturally be impossible. On the other hand, nomic connection has nothing to do with the feature of intentionality that does not allow substitution of co-referential terms. It is perfectly sensible that a subject has a belief "that *s* is *F* without believing that *s* is *G*, despite the nomic connection between *F* and *G*".²⁷² In other words, if one replaces the belief content that *s* is *F* with the other belief content that *s* is *G*, even though nothing can be *F* without being *G*, then he changes the truth value of a belief content provided the fact that the subject holding the belief that *s* is *F* does not believe that *s* is also *G*.

These two issues are very important issues for the informational account. Let us now have a look at Dretske's treatment of the second issue. It is clear from the second issue that there appears to be an insurmountable difficulty about the asymmetry between the way informational theory explains the substitution of co-referentially equivalent pieces of information and the way folk-psychology explains the substitution of co-referentially equivalent contents of propositional attitudes. While the former relates the possibility of substitution to the presence of nomic connections between two events, the latter simply denies the role of nomic connections for the possibility of substitution of co-referentially equivalent contents. Without explaining this asymmetry,

²⁷² Ibid., p. 62.

Dretske believes, informational theory could not be said to capture “the full intentionality of beliefs”.²⁷³

But we still have not captured the full intentionality of beliefs. In teaching our child the concept *water*, for instance, why say that the structure that develops to encode information about water is not, instead, a structure that was developed to encode information about the presence of oxygen atoms? After all, any incoming signal that carries the information that *s* is water carries (nested in it) the information that *s* has oxygen atoms in it (since there is a lawful regularity between something's being water and its having oxygen atoms in it).²⁷⁴

Dretske makes an interesting move here. In order to present his philosophical maneuver about this issue, I need to take you on a brief tour into the discussion of sensory and cognitive representation, that is, the difference between seeing an *s* which is *F* and seeing that *s* is *F*. Only when these two forms of representation are understood does Dretske's solution to the relevant question become more explicit. So, let us begin with the discussion of information in analog and digital forms.

Information-processing models in the literature, from Dretske's point of view, hold the idea that perception is an activity of systems that stands for the reception and delivery of information, and cognition stands for the utilization of that information. In this information-processing framework, perception and cognition mean only different degrees of the information-processing operations. In other words, there is a certain line of continuity between perception and cognition ranking from the low-degrees of the information-processing activity, which stands for the perceptual states of a system, to the high-degrees of the information-processing activity, which stands for the cognitive states of a

²⁷³ Ibid., p. 62.

²⁷⁴ Ibid., p. 62. Italics belong to Dretske.

system. Hence is the reason for the motto: perceptual states are just low-grade forms of cognitive states.²⁷⁵

This is one of the points where Dretske's account of information vividly manifests a strong disagreement with the earlier information-processing models. For him, in interpreting perceptual states such as seeing, hearing, etc., as low-grade forms of cognitive states one does nothing other than assimilation of the distinctive features of sensory experience to cognitive states and events. In order to break out from this circle of assimilative modeling, Dretske proposes a model of information which fairly explains different ways incoming information are being encoded with respect to the function of perceptual and cognitive centers in the entire economy—efficient use of the resources—of information-processing systems.²⁷⁶

According to the traditional sense of the difference between analog and digital encoding of information, one can define analog coding as “a continuous representation of some variable property at the source” and digital coding as “a discrete representation of some variable property at the source”.²⁷⁷ Take, for instance, the speedometer on a car's dashboard that basically consists of numbers from 0 to 100 and a pointer designed to indicate one of those numbers in accordance with the informational measure of the speed of a car. Each different position of the pointer

²⁷⁵ Dretske, *Knowledge and The Flow Of Information*, p. 135. It seems to me that these models borrowed the old idea from British empiricist philosophy in which ideas are considered to be less lively copies of sense-impressions. At least, as far as I can observe, there is a common point between the information-processing models and British empiricist philosophy with respect to the claim about the continuity between perception and thought.

²⁷⁶ Dretske, *Knowledge and The Flow Of Information*, p. 135.

²⁷⁷ *Ibid.*, pp. 135-6.

represents a different speed state/value of the car. This is why this mechanism is a continuous representation of some variable property at the source and therefore an analog encoding mechanism. As for the digital coding, let us have a look at the example of the light on a dashboard designed to indicate oil pressure of a car. This device occupies an active state (say, constant flashing of red light in a box on the dashboard) whenever registering low oil and it occupies a passive state (not turning red light) whenever there is no registry of low oil. In this oil-pressure measure device, there are “two informationally relevant states,”²⁷⁸ either it is on or off. This device presents only two states that are discrete and thus no continuity between them. Thus, an oil-pressure measure device is a digital device.

Dretske wants to make some slight changes in the traditional explanation of digital and analog encoding of information. In what follows, Dretske presents his own way of understanding this distinction. His distinction shows “the different way facts can be represented”.²⁷⁹

Definition of information in digital form: “[A] signal (structure, event, state) carries the information that *s* is *F* in digital form if and only if the signal carries no additional information about *s*, no information that is not already nested in *s*’s being *F*.”²⁸⁰

Definition of information in analog form: “A signal (structure, event, state) carries the information that *s* is *F* in analog form iff “the signal does carry additional information about *s*, information that is not nested in *s*’s being *F*”.”²⁸¹

²⁷⁸ Ibid., p. 136.

²⁷⁹ Ibid., pp. 136-7.

²⁸⁰ Ibid., p. 137.

²⁸¹ Ibid., p. 137.

In order to understand the reason behind these two definitions, let us review the example given by Dretske. There is obviously a very sharp difference between a picture and a statement with respect to their information-carrying functions. Think of a statement “the cup has coffee in it” and think of a picture of the cup that has coffee in it. In the statement, there is not any additional information given about the cup. It has coffee in it, that’s all we get. In contrast, the picture of the cup that has coffee in it gives you much more information than only the state of the cup filled with coffee. For instance, it automatically gives you information about the shape of the cup, the colour of the cup, etc. These sorts of information, in Dretske’s terminology, are not nested in *s*’s being *F*.²⁸² The most important point about these two examples is that they represent the same fact; e.g., *s* is *F*, in different ways. This is very crucial point in Dretske’s agenda.

Furthermore, this example reveals the fact that in the analog form of information we have various details of data about the source, whereas in the digital form of information what we have is very specific information about the source. This is again very important for Dretske. According to him, analog encoding of information and digital encoding of information help us understand sensory and cognitive processes. In Dretske’s understanding of this relationship, our sense organs, eyes, ears, touch, etc., all provide information in analog form. There are lots of

²⁸² Ibid., p. 137. See also pp. 70-75 for Dretske’s explanation of the nesting relation and its implications on the informational content of a signal. Of course one can argue that it is possible to translate the informational elements of the picture into a statement describing, albeit in a very long clause, those elements. But, this does not change the fact that one provides information in analog form and the other provides information in digital form.

incoming perceptual details about our surroundings, most probably we are not really aware of them all, yet they are all stored in the perceptual centers in analog form.²⁸³ One can then think of a digital-conversion unit converting the rich and messy information stored in the analog device of a car, like a simple speedometer, into some very specified range of information for the utility of car and driver. In the same manner, Dretske conceives, we have a digital-conversion unit in our mind whose function is to extract “pertinent information from the sensory representation”²⁸⁴ in order to create a conceptual space (a space of concepts) for the cognitive activity such as beliefs, thoughts, etc. For Dretske, “it is the successful conversion of information into (appropriate) digital form that constitutes the essence of cognitive activity”.²⁸⁵

This framework does not only explain the nature of sensory and cognitive processes, it also allows for the possibility that within a sensory process, we can have a sensory representation of the information that *s* is *F* without having a cognitive representation of the fact that *s* is *F*. Put differently, one can perceive something, *s*, which is *F*, without having any belief that *s* is *F*.²⁸⁶ This is a very comforting feature of Dretske’s theory,

²⁸³ This is very similar to Ned Block’s distinction between phenomenal consciousness and access consciousness. See Ned Block, “On a Confusion about the Function of Consciousness,” *Behavioral and Brain Sciences* 18, 2 (1995), pp. 227-247.

²⁸⁴ Dretske, *Knowledge and The Flow Of Information*, p. 142.

²⁸⁵ Ibid., p. 142. In a footnote to this statement, Dretske, on p. 254, cautions the reader against thinking that the condition of conversion of information from analog to digital form is sufficient for the attribution of perceptual-cognitive states to any system having a digital-conversion unit. For this reason, one should not think as if a simple measuring device with a digital-conversion unit has a perception or knowledge about the event at the source. According to Dretske, the capacity to convert analog information to digital information is only a necessary condition for a system to be a genuine perceptual system.

²⁸⁶ Ibid., p. 142.

because it explains the traditional idea that knowledge, belief and thought always presuppose concepts while sensory experience does not presuppose concepts. Thus, it is plausible to interpret Dretske's distinction between analog and digital forms of information as sensory and cognitive representations in the old traditional thought.

If perception is understood as a creature's experience of his surroundings, then, perception itself is cognitively neutral. Nevertheless, although one can see (hear, etc.) an *s* which is *F* (sensorily encode information about *s* and, in particular, the information that *s* is *F*) without believing or knowing that it is *F* (without even having the concepts requisite to such beliefs), perception itself depends on there *being* a cognitive mechanism able to utilize the information contained in the sensory representation. In this sense, a system that cannot know cannot see; but if the system is capable of knowing, if it has the requisite cognitive mechanisms, then it can see without knowing.²⁸⁷

This is one of the passages indicating a very interesting modification in Dretske's earlier idea in *Seeing and Knowing*.²⁸⁸ In *Seeing and Knowing*, Dretske also emphasized the idea that perception of *X* is independent of knowing *X* and here in KFI he holds on to the same idea. It is very remarkable here in KFI that he also adds a very important requirement: only if a subject has the capacity to digitalize the analog information provided by sensory processes for the utility of cognitive centers is it possible for him to see something without knowing it. This modified thought definitely contributes a new dimension to his earlier thought by adding a new flavor (of holism). Now, it is not enough, although very important, that sensory representation and cognitive representation are distinguished with respect to their individual roles in a system. A system

²⁸⁷ Ibid., p. 153. Italics are Dretske's.

²⁸⁸ Fred I. Dretske, *Seeing and Knowing* (London: Routledge & K. Paul, 1969).

may have a certain structure consisting of sensors to collect information about its environment in analog form, but as long as “this structure” does not have “a certain function within the larger information-processing enterprise” by making that information in analog form “available to a suitable converter for possible cognitive utilization,”²⁸⁹ it will not be able to see anything. Therefore, for any system to be qualified as having perceptual states, it is required that a structure (in that system) containing analog information must be connected to a cognitive mechanism that is capable of digitalizing that given analog content; i.e., it must be able to convert given sensory representations to a specified, classified, recognized cognitive representation.²⁹⁰

As a summary, Dretske begins with the idea that sensory processes must be distinguished from cognitive processes. Then, he identifies perceptual experience with information processing in analog form. This analog information, then, must be available to a digital-conversion unit of cognitive centers. Analog-digital transformation, or sensory-to-cognitive representation conversion, is a process of abstracting the concrete, generalizing the particular information. Without this mechanism being present and operational, one cannot talk about recognition, classification, categorization, and experience, in short any cognitive or perceptual activity at all.

Now, given the requisite background about Dretske’s distinction between sensory and cognitive representation, it will be appropriate to turn back to the second issue and examine how Dretske handles the

²⁸⁹ Dretske, *Knowledge and The Flow Of Information*, p. 153.

²⁹⁰ Ibid., p. 256, footnote 29.

matter. The problem was that there is an asymmetry between the ways informational theory and folk psychology explain the reason for the failure of substitution. Why is it the case in relatively primitive information-processors that if they carry the information that something is water, then they also carry the information that that thing has oxygen atoms in it—since there is a nomic connection between something's being water and that thing's having oxygen atoms in it—but the same does not hold for human beings? After all, if informational theory addresses the fact that human beings are no different from other information processing systems with respect to the capacity of information production and consumption, then it needs to point out the cause of the explanatory gap between primitive information processors and sophisticated information processors like human beings; it needs to find a way out of this apparent explanatory gap.

By making use of Dretske's distinction between sensory and cognitive representation and especially his thesis that without a cognitive mechanism nothing can see anything, we can draw his line of reasoning as follows. Let us take beliefs as the paradigm case of cognitive activity. Belief requires concepts. That is to say, without having a concept about *s* and *F*, it is not possible to have a belief that *s* is *F*. However, experience does not require concepts. That is to say, without having a concept about *s* and *F*, it is still possible to experience an *s*, which is *F*. Experience "itself depends on there being a cognitive mechanism able to utilize the information contained in the sensory representation."²⁹¹ In order to have an experience of anything, it is required for a system/being to be able to

²⁹¹ Ibid., p. 153.

convert information in analog form into information in digital form. That is to say (in Dretske's terminology), without any cognitive mechanism like analog-to-digital conversion units that can convert sensory representations to a specified, classified, recognized cognitive representation, no system/being will be capable of experiencing anything at all. Without cognitive capacities, no experience is possible. Remember the Xerox Principle, which reads that if there is a nomic regularity between Fs and Gs such that all F's are G, then any signal carrying the information that s is F necessarily carries the information that s is G as well. Now, consider that there are two systems/beings, A (like television, radio, etc.,) and B (like vervet monkeys, human beings) that differ from each other with respect to their information-carrying structures. Suppose that A has an information-carrying structure that can encode information only in analog form and cannot utilize that information for further purposes in cognitive centers due to the absence of an analog-to-digital conversion unit; and B has an information-carrying structure that can encode information both in analog and digital forms by its recording sensors and analog-to-digital conversion units. For simplicity, let us suppose that A's information-carrying structure consists of an analog machine and B's information-carrying structure consists of analog and digital machines plus a converter machine for converting analog information into digital information. Let A and B both receive a signal carrying the information that s is F, and let there be a lawful regularity between s's being F and s's being G. In other words, nothing can be F without being G. (Remember Dretske's definition of information in analog form) A signal (structure, event, state) carries the information that

s is F in analog form iff “the signal does carry additional information about s, information that is not nested in s’s being F”.²⁹² If A has the information processor that can encode information only in analog form, i.e., if A has only an analog machine, and if A receives a signal carrying information that s is F, then it is necessary for A that it also receives the information that s is G (because of the Xerox Principle and definition of information in analog form). Therefore, A also necessarily receives the information that s is G. (Now remember the definition of information in digital form) A signal (structure, event, state) carries the information that s is F in digital form iff the signal carries no additional information about s, no information that is not already nested in s’s being F. If B has the information processor/structure that can utilize the information in analog form contained in sensory representation for further purposes in cognitive centers by its analog-to-digital conversion mechanism, and if B receives a signal carrying the information that s is F, it is possible for B to have a belief that s is F without having the belief that s is G (even though there is a lawful regularity between s’s being F and s’s being G). In Dretske’s terminology, B may experience something, s, which is F and G at the sensory level, but not yet able to identify that nothing can be F without being G at the cognitive level because B may not yet develop a conceptual sensitivity to G-eliciting stimuli as it does to F-eliciting stimuli. Therefore, it is possible for B to have a belief that s is F without having the belief that s is G even though it may experience s, which is

²⁹² Ibid., p. 137.

both F and G. In other words, it is possible for B not to exploit the information that s is G in digital form.²⁹³

The gist of the preceding argument is all about showing the difference between primitive and relatively pedestrian processing mechanisms of beings/systems and sophisticated and relatively more resourceful processing units of beings/systems. By this demarcation, now there is no problem, Dretske maintains, about the explanation of the fact of asymmetry between the way the informational account explains the possibility of substitution of co-referentially equivalent pieces of information and the way folk psychology explains the substitution of co-referentially equivalent contents of propositional attitudes. If the object of inquiry is a primitive information-processor where information can be encoded only in analog form, and if it receives a signal carrying the message that s is F, and if Fs are all correlated with Gs in accordance with a nomic regularity, then it will necessarily carry the information that s is G. But, if the object of inquiry is a sophisticated information-processor where information can be encoded both in analog and digital forms, and if it receives a signal carrying the message that s is F, and if there is no F without being G, then it is still possible that it may have sensory representation of F and G at the sensory level, but not have cognitive representation of G while it cognitively represents F, or vice versa.

I will not get into the details about the problems with his explanation, but only remark on two initial difficulties. First, Dretske introduces a distinction between experience and belief in order to explain the asymmetry between primitive information systems and sophisticated

²⁹³ This is my reconstruction of Dretske's lines of reasoning.

information systems (e.g., human beings). But, we do not really see an argument that supports this distinction; he only takes this distinction to be obvious and tells a story about the different ways sensory centers and cognitive centers of an information system encode any given piece of information. Contrary to Dretske's view, one can find arguments that support the view that we cannot really have experience of something if we do not have a concept to identify that object experienced. So, without having any argument, I seriously doubt about the success of Dretske's story in explaining the problem about the substitution of co-referentially equivalent pieces of information. Secondly, Dretske's story about analog and digital encoding of information really does not work at all. Let's turn back to the case in point. Dretske says: "In teaching our child the concept *water*, for instance, why say that the structure that develops to encode information about water is not, instead, a structure that was developed to encode information about the presence of oxygen atoms?"²⁹⁴ Let's just focus on the case of the child, and how this contrasts with Dretske's project of naturalizing the mind. Exactly how can teaching a concept "water" be explained with Dretske's distinction between the analog and digital ways information can be encoded? Should we say, following Dretske, when a child looks at something, which happens to be water and having oxygen atoms in it, she has a sensory representation of both water and oxygen (all in the analog form), but when a mother teaches her the concept "water", then the child digitalizes all the relevant information in perception as of being one kind, e.g., water? Most importantly, how does a mother come to teach the concept of water in the first place? If the

²⁹⁴ Dretske, "Précis of Knowledge and the Flow of Information," p. 62.

target of the project of naturalization, as Dretske says, is never to mention intentions, social conventions, etc., then it seems to me that the project fails even right at the beginning here. Learning a concept of water is more than just processing of information in a digital form, it has to do with being a member of society where “water” is important for the individuals and the concept of water has been dubbed the way it is in a society.

Let us now move on to examine Dretske’s treatment of the misrepresentation issue. According to his view, we should have a good comprehension about information-processing units and mechanisms, especially how they acquire information-carrying roles/functions in a system of which they are a part. Consider maps. Let us say that there is a strict and steady convention among mapmakers such that blue ink means “water”. Imagine that there are plenty of conventional agreements on how to use many other symbols. So, we have a good criterion upon which we can decide whether a given map is an accurate representation of a specific location. For instance, if one puts a blue ink on a certain region of the map and if there is no water in the place about which the map is to give information, then we have a case of misrepresentation, because blue ink, according to the common convention, acquires a particular information-carrying role and when it fails to perform that role, we will have a misrepresentation of certain facts about a certain location.²⁹⁵

Following this thread of thought, Dretske believes that one can think that information-bearing structures, for instance neural structures

²⁹⁵ Ibid., p. 62.

in human beings, acquire information-carrying roles during their development with experience and learning through which those structures *naturally*²⁹⁶ develop a responsiveness to a particular kind of external stimuli carrying a definite piece of information. Consider how we acquire a concept of bird for instance. While learning what a bird is, we are typically exposed “to positive and negative instances of the concept” of bird in order to help us gain “a sensitivity to the kind of information (that s is a bird) that defines the concept”.²⁹⁷ After the termination of the learning period during which we are tested to distinguish birdy phenomena from non-birdy phenomena, according to Dretske, we succeed in acquiring an information-carrying role for some neural structures in our head that consistently respond to the kind of information that something is a bird.²⁹⁸ But then again, it is still possible for the student in the post-training period to mistake an airplane for a bird. He might have never come across any airplane during the learning phase. So, he may perceive an airplane in the post-training period and utter the word “bird” thinking that the airplane instantiates one of the most essential dispositions of being a bird, “ability to fly”. Dretske believes that we have here a case of misrepresentation. For neural structures, which acquired an information-carrying role for a relevant

²⁹⁶ According to Dretske, here we have a case of neural structures that do not acquire an information-carrying role by conventional assignment, but rather by natural development of structures during learning and experience. That’s why I used italics here in order to emphasize this point. On this point, see Dretske, “Précis of Knowledge and the Flow of Information,” p. 62. Although Dretske believes that he presents an account for natural representation as opposed to acquired / conventional representation, Fodor and Cummins particularly emphasize that his view can be used to explain only learned concepts and representations, so Fodor and Cummins believe that no account is actually given for mental / natural representations.

²⁹⁷ Dretske, “Précis of Knowledge and the Flow of Information,” p. 62.

²⁹⁸ Ibid., p. 62.

piece of information (that something is a bird), are obviously triggered by the external stimulus/source that does not generate the information (that something is a bird). Then, Dretske says, those neural structures fail to perform what they have learned to do during the learning phase.²⁹⁹ Mere nomic covariation is not enough; we must introduce information-carrying roles/functions of the receivers in order to give an account for the cases of misrepresentation between the receivers and sources of information.

Let's get into some technical details about Dretske's account here. I think that it will help us understand the source of the problem. In Grice's natural meaning, whenever we confront cases where X means Y, this "X means Y" entails Y. In Dretske's account of information/indication, whenever there is an informational relationship between two types of events, say a thunder sound and the lightning channel that produced it, and since information is something capable of yielding knowledge and therefore requires truth, then this entails the truth about the existence of that lightning channel. This approach, as I noted before, gives rise to a very serious problem in Dretske's account. Now, let us illustrate the problem. Let us suppose that a subject has the sensory representation of the fact that s is F. There are great varieties of sensory representations in our experience of s, some of which are Fs and some of which are Gs, or in our experience of something else (say t), some of which are Fs and some of which are Gs. At this stage, we will perceive s, which is F, but not able to know whether s is F. According to Dretske, there is a digital-conversion unit that converts the incoming information in analog form

²⁹⁹ Ibid., p. 62.

into digital form by categorizing and classifying it under some selective method. Then, at the end of this process, we have a cognitive representation of the fact that *s* is *F*; we know that *s* is *F*. As a result, if our sensory and cognitive mechanisms are in perfect order, then we will have the belief-content that *s* is *F*. But, it is certainly true that we have also lots of false contents in our mind. We can misrepresent certain situations in the world. For instance, we can misrepresent a particular situation, where *s* is *G* or *s* does not even exist in that situation, and say that *s* is *F*.³⁰⁰ Then, if there is nothing like misinformation/misindication as a kind of information/indication, then how can Dretske's account explain the cases of misrepresentation? It seems that as long as the sensory and cognitive mechanisms are operational and in order, then we will always have the correct content about things outside; yet even though there can be nothing wrong with our mind, we can still have false beliefs. Dretske aims to reduce semantic content, the content we have while believing, hoping, desiring, etc., to the informational content, and explain intentionality and intelligence via the informational account. But, without giving an account for the cases of misrepresentation, Dretske's project is doomed to fail.

In Dretske's account, there must be a nomic dependency between two types of events/states of affairs in order to establish a relation of information between these two points. The dance of a honeybee is

³⁰⁰ For Dretske, there are two kinds of misrepresentation. In the first kind, there is an object, but the object the system, say *S*, represents to be *F*, say blue, is not *F*, blue. In the second kind, *S* again represents something to be blue, but in fact there is no object that *S* represents to be blue. For an extensive discussion of these two kinds of misrepresentation, see especially Fred I. Dretske, *Naturalizing the Mind* (Cambridge, Mass.: MIT Press, c1995), pp. 26-27.

informationally valuable for other members of the honeybee group because each dance of a honeybee is tokened when and only when there is nectar nearby. But, one may argue for the fact that it is still possible for some honeybees, due to extrinsic or intrinsic factors, to act out this ceremony about the nectar when there is really no nectar around. The dance of that honeybee will still have the content “there is nectar nearby” even though it is not tokened by the right cause.

Following the insight from conventional functions for determining representational content and misrepresentation, Dretske then claims that ontogenetic function acquisition and learning must be added to the causal story about mental representations. Let us focus again on concept acquisition. Suppose that there is a student, S, and his trainer/teacher helps him learn and correctly apply simple concepts such as the concept of robin, raven, bird, elephant, etc. S perceives objects such as robin, raven, elephant, dog, etc., and his trainer/teacher teaches him the correct concepts corresponding to objects in his visual field. The teacher also examines whether S really learns those concepts by examining whether he can employ them in the right circumstances. After the training period is over, S is given the license to use those concepts. He graduates from the school. It is still possible, Dretske underlines, for S to use, in his box of concepts, a concept that is not triggered by the right cause, but only this time it will be regarded as wild tokening of that concept. Since S is, in the post-training period, a person who knows all about the concepts he has acquired, those tokenings accidentally triggering the concept are now counted as wild and will not sneak into the representational content of that concept. For instance, if the student in the post-training period

mistakes a fox for a dog, then his concept “dog” will not mean “dog or fox”; it will still mean/represent dogs. In a nutshell, Dretske makes a distinction between what happens in the training period and what happens afterwards, and by means of this distinction, he claims that wild tokenings of a concept can only occur after the student knows everything there is about that concept. Those wild tokenings are delicately differentiated from the normal tokenings of a concept. So, when a student, after the training period, utters a sentence “This is F” in the absence of the relevant object or in the presence of objects that are G, H, etc., but not F, this will be an instance for him to misrepresent facts. The problem of misrepresentation is solved and the case is closed, so says Dretske.³⁰¹

However, there are serious difficulties in his solution. First of all, the training period example makes the assumption as if there is a very sharp, principled ground where the learning definitely comes to an end and the student has a full command of the concepts that he has learned in the training period. This seems to be a very incredible assumption. It is much more reasonable to hold a modest assumption that learning concepts is a lifetime activity³⁰² and there is no such point where one can say “Behold, I am now the ultimate authority on the correct application of the concept in question”.

Secondly, we should pay attention to the implicit assumptions in the example. The example implicitly appeals to many things such as the

³⁰¹ Dretske, “Précis of Knowledge and the Flow of Information,” p. 62.

³⁰² Robert Cummins, *Meaning and Mental Representation* (Cambridge, London: MIT Press, 1989), p. 68; Jerry Fodor, “Semantics, Wisconsin Style.” In *A Theory of Content and Other Essays* (Cambridge, Mass.: MIT Press, c1992), p. 41.

intentions of the teacher, going to school, use of the concepts in accordance with the conventions of the society, etc., things that violate the requirements of naturalism, according to which any mention of things like intentions, conventions, etc., must be avoided during the explanation of mind and representation. A quick look at the example, however, will bring all these things to the surface. During the training period, a subject is trained to use "X" as a mental representation of birds. Dretske says that when this student graduates from "X" teaching school, and mistakes an airplane for a bird, then we would say that airplane is not one of the normal causes for "X". So, whenever any airplane triggers "X", then we would say that it is false tokening of "X" and we have a case of misrepresentation. But, it is possible to interpret the same situation counterfactually. If any airplane can cause "X" in the post-training period, then it could have been true of that airplane to cause "X" in the training period as well. Just imagine one day before the graduation the student looks at the sky; he sees an airplane and mistakes it for a bird. So, instead of saying that the student learnt "X" means bird, we can equally say that the student might have learnt "X" means airplane or bird. Then, when the student in the post-training period utters the word "X" in the presence of airplane, we should say that there is no misrepresentation; because he had previously learnt that "X" means airplane or bird. In Dretske's example, however, you can feel that the teacher's intentions, conventions of the society, schooling, etc., play a central role in determining how and why the student learns that "X" means bird rather than "X" means bird or airplane. So, when you emphasize, in the example, the role of the teacher, schooling, conventions of the society

about the meaning of a concept, you may get around the problem at the cost of failure of the project of naturalism, which is not a good trade.³⁰³

Naturalism is one of the big goals in Dretske's theory, so I do not think that he will ever consider abandoning it seriously. In a nutshell, I believe that Dretske's story about learning is directly in conflict with his big project of naturalization.

Thirdly, and most importantly, telling a nice story about the occurrence of misrepresentation does not seem to help Dretske deal with the problem of giving an account for the possibility of misrepresentation. Personally, I think that the learning story is nice, and it just explains how misrepresentation can occur. But, I do not think that telling a story about the occurrence of misrepresentation can help overcome the difficulty with Dretske's account. I shall briefly explain this point. Dretske's account straightforwardly implies that if a signal carries the information that *s* is *F* then *s* really is *F*. So, on this account, when my mental state represents that *s* is *F*, then it must be the case that *s* is really *F*. There is no way my mental state can misrepresent a certain situation. Dretske's theory already rules out misrepresentation, and it just does not help a bit to tell a story how misrepresentation can occur. Saying something about how misrepresentation can occur and solving the problem of misrepresentation in the theory are two different things.³⁰⁴

In this chapter, my primary goal was to understand and present Dretske's informational account of representation as it was first

³⁰³ Jerry Fodor, "Semantics, Wisconsin Style," pp. 39-41; Adams, Fred and Aizawa, Ken. Spring 2010. Causal Theories of Mental Content. Available [online]: "<http://plato.stanford.edu/archives/spr2010/entries/content-causal/>".

³⁰⁴ I owe this point to Stephen Voss.

advertised in KFI.³⁰⁵ I began with his technical formulation of the concept of information, then I moved on to his distinction between sensory and cognitive representation. In the end of this chapter, I talked about two important problems (substitution of equivalent pieces of information and misrepresentation) in his account and how he deals with those issues. It seems to me that none of Dretske's solutions really work. I tried to explain some of the issues with his solutions. In the midst of all these issues, I think that there might be only one thing that Dretske understands well: mere nomic correlation between the receiver (mind) and the source (world) is really not enough to ground a representation-relationship.

³⁰⁵ The informational account has always been present in Dretske's works; however, later on in his career he added a teleosemantic component to this informational account. Cf. Fred I. Dretske, *Explaining Behavior: Reasons in a world of causes* (Cambridge, Mass.: MIT Press, c1988); Fred I. Dretske, *Naturalizing the Mind* (Cambridge, Mass.: MIT Press, c1995).

CHAPTER VI

MILLIKAN'S ACCOUNT OF REPRESENTATION

Dretske's account in *Knowledge and the Flow of Information*³⁰⁶ was probably one of the important milestones in the twentieth century philosophy of mind. Not only did it offer a very fertile perspective through which the notion of representation is understood in terms of information/indication, it also did contribute to the field of artificial intelligence, in the interim, by creating a model of cognition and intelligence that is compatible with the broadly optimistic research projects and predictions about functional networks and computational structures in that field since Alan Turing's notorious works. By reducing representation to information/indication, Dretske particularly emphasized the fact that any item can be a representation of something else without any subjective-interpretive activities around. Finally, his account provided a very naturalist position, according to which there is no difference, in principle, between structurally primitive information processors, like television sets, and relatively more sophisticated information-carrying structures, like perceptual and cognitive mechanisms of vervet monkeys, human beings, robots, etc. More precisely, it proved to be one of the fine defenses of a naturalist-functional position, according to which it simply becomes possible to begin with a very simple information mechanism and, given requisite materials and conditions, build a very complex intelligent being/system

³⁰⁶ Fred I. Dretske, *Knowledge and The Flow Of Information* (Stanford, CA: CSLI Publications, 1999; originally published: Cambridge, Mass.: MIT Press, c 1981).

at the final phase of the assembly line. This was nothing but a great salvation from the Cartesian intuitions, the must-step in naturalizing the mind.³⁰⁷

There were, however, very dubious sides of Dretske's indication approach in *KFI*. First of all, it seemed to many people, including later Dretske, that it cannot solve the problem of misrepresentation. Furthermore, it would be quite problematic for this theory to explain our capacity to represent non-existent things, if the representing-relation is nothing but a reliable correlation between an inner state and a state of affairs outside. The philosophical climate was not comforting for Dretske at all. David Papineau and Ruth Millikan, in the middle of this hot climate, independently developed a theory of content/mental representation called "teleosemantics"³⁰⁸ which, according to their opinion, achieves much more fruitful results on the test about Brentano's criterion of intentionality than Dretske's account. In this chapter, by extracting Millikan's views about conditions of representation, I aim to present at least one of the most powerful versions of teleosemantic theories.

³⁰⁷ The goal of naturalizing the mind has been one of the most important themes in Dretske's later writings as well. See also Fred I. Dretske, *Explaining Behavior: Reasons in a world of causes* (Cambridge, Mass.: MIT Press, c1988); Fred I. Dretske, *Naturalizing the Mind* (Cambridge, Mass.: MIT Press, c1995).

³⁰⁸ Alternatively, it is also called a teleonomic or teleological theory of content. In this chapter, I will present Millikan's version of teleosemantics. For Papineau's version of teleosemantics, see David Papineau, "Representation and Explanation," *Philosophy of Science* 51, 4 (December, 1984), pp. 550-572; David Papineau, *Reality and Representation* (Oxford UK, Cambridge USA: Blackwell, 1987); David Papineau, *Philosophical Naturalism* (Oxford UK, Cambridge USA: Blackwell, 1993). For an extensive discussion of teleosemantics, see Graham Macdonald and David Papineau (Eds.), *Teleosemantics* (Oxford, UK: Clarendon Press, 2006).

Any theory of content/mental representation, among other things, must be able to tell us why a subject having a thought about birds singing in the garden of his house has that thought rather than some other thought about the recent electric power failures in the neighborhood or the unbearable lightness of being or nothing at all.³⁰⁹ Aristotle's theory of mind tells us that it is because of the causal relation between the objects of sensation/thought and our sense organs/intellect, that our inner states gradually resemble their objects with which they go into perceptual and intellectual interaction. That's why we have a mental state with the content p. On the other hand, Dretske's informational theory, denying similarity as the ground of representational relations, tells us that it is because of an informational bridge, plus some other conditions, between the source and receiver that we have a mental state with the content p. Mental representations, provided that the mind falls within the extension of nature, i.e., if it is a part of nature, are, from his point of view, like natural signs through which information is carried from one point to another point.

For Millikan, the basic causal or informational theories of content are in serious and unsolvable trouble. If representational activity, as causal theory has it, is all about a (detection) mechanism, trait or process whose function is to produce inner states, "that correspond to or covary with something in the outside world",³¹⁰ then the doom of the causal theory is inevitable. For not every device/mechanism whose job

³⁰⁹ Neander, Karen. Spring 2012. Teleological Theories of Mental Content. Available [online]: "<http://plato.stanford.edu/archives/spr2012/entries/content-teleological/> [Spring 2012]".

³¹⁰ Ruth Millikan, "Biosemantics," *Journal of Philosophy* 86, no. 6 (June 1989), p. 283.

description covers production of inner states that covary with the states of affairs in the external world is a representation producer device/mechanism.³¹¹ For instance, the pigment arrangers in the skin of a chameleon have the function to covary the chameleon's skin color with the most immediate object in its environment, but, according to Millikan's perspective, this is not a case of representation production.³¹² Then, the basic causal theory does not really give us any tool to distinguish mechanisms producing representations from mechanisms not producing representations.

On the other hand, if we hold the claim that "*having the function of representing R*",³¹³ as Dretske and Stampe³¹⁴ suggested, is nothing beyond "a natural sign or representer of R when the system functions normally",³¹⁵ namely if representational activity of an inner state is all about its being a natural sign between two things so that one is (naturally) carrying information about the other system whenever the system is in a good order, then we could be justified in asserting that this approach, at least, enables us to delimit the boundaries of informationally relevant states of mechanisms so that not all covarying states of those mechanisms are counted under the class of representations. Unfortunately, even this approach, for Millikan's standards, cannot give us any good clue about what representation is. First of all, "the

³¹¹ Ibid., p. 283.

³¹² Ibid., p. 283.

³¹³ Ibid., p. 283.

³¹⁴ Dennis W. Stampe, "Towards a Causal Theory of Linguistic Representation," *Midwest Studies in Philosophy* 2, no. 1 (September 1977), pp. 42-63.

³¹⁵ Millikan, "Biosemantics," p. 283.

production of natural signs is undoubtedly an accidental side effect of normal operation of many systems".³¹⁶ One can take a red face as the natural sign of the fact that one is exerting himself, or that he has been in the heat or that he was burned.³¹⁷ As the same goes for what the heart does like production of a thumping sound as a natural sign, the by-products of a mechanism—no matter how often it takes place—is not the defining characteristics or the real function of that mechanism.³¹⁸ But, even though we try to be charitable in our judgment and accept that representations can be by-products of mechanisms, still more damaging an issue remains intact. For, it is clearly obvious that it still does not follow that "representations must carry natural information".³¹⁹ Let's suppose, as Dretske would have it, that the chameleon's skin color changes, because the pigment arrangers, under normal conditions, are supposed to signal danger. In Dretske's terminology, we express this as follows: there is a mechanism whose job / function is to indicate or carry information about the threat in the external world. But, if there is an informational relationship between a source, *s*, and a receiver, *r*, then *s* must exist. However, in many instances, we clearly observe that nature cares more for the survival value of danger signals, and less for the truth-value whether those signals correspond to real threats outside. It is always "better to err on the side of caution".³²⁰ For this reason, "there is

³¹⁶ Ibid., p. 283.

³¹⁷ Ibid., p. 283.

³¹⁸ Ibid., p. 283.

³¹⁹ Ibid., p. 283.

³²⁰ Ibid., p. 283.

nothing incoherent in the idea that" it is natural for mechanisms whose (proper) functions are to produce signals not to "carry natural information concerning the dangers they signal".³²¹ Then, it follows that there is actually no constraint that representations, even when supposing (for the sake of discussion) that natural signs could be taken as representations, must carry natural information about their source.

Adopting Millikan's line of thought is, in a sense, very revolutionary and revisionary to what people take to be true in accordance with the basic causal and informational theories. According to her point of view, mental representation has nothing to do with natural signs, nor with the function of carrying information, much less with covarying states. In order to grasp Millikan's revolutionary and revisionary perspective, it is extremely decisive, at this stage, to make room for Millikan's terminology in her toolbox. For Millikan, any representational relation arises from a proper function of consumer mechanisms that utilize that representational relation for the system.³²² Ordinarily speaking, from the term "proper function" of X we understand what X is supposed to do, what it has been designed to do, or what it ought to do. But, what does Millikan understand by a proper function?

An item X has proper function F only if (i) X is a reproduction of some prior item that, because of the possession of certain reproduced properties, actually performed F in the past, and X exists because of this performance; or (ii) X is the product of a

³²¹ Ibid., p. 283.

³²² In some other versions of teleosemantics, it is dependent on production mechanisms rather than consumer mechanisms.

device that had the performance of F as a proper function and normally performs F by way of producing an item like X.³²³

In order to explain content-fixing conditions and thereby distinguish true representations from false representations, both causal and informational theories explicitly appeal to the idea that a causal or dispositional feature of an item signifies what it does in general or what it is disposed to do provided that certain conditions obtain.³²⁴ Consider Dretske's informational theory: there is a mechanism X which is disposed to represent (indicate or carry information) R that s is F, given that (i) s is F, and (ii) X runs normally, and (iii) there is a law-like regularity between s's being F and X's being disposed to do whatever it does/production of R.³²⁵ Millikan doubts whether this is a good way to understand the nature of representation. These sorts of causal or dispositional analyses, at the end of the day, only tells us what something does in general or what it is disposed to do in accordance with a nomic regularity; but what something does in general might not be its proper function, what it has been designed to do. In other words, these sorts of analysis methods only give us statistical facts, averages or commonalities,³²⁶ and the proper function of something may not be collapsed to what it does frequently. Consider three cases: (1) X may realize many things, "not all of which are

³²³ This is Rowland's reconstruction of Millikan's definition of a proper function. See Rowlands, Mark. *Teleosemantics*. Available [online]: "<http://host.uniroma3.it/progetti/kant/field/teleo.html>". See also Ruth Millikan, "In Defense of Proper Functions," *Philosophy of Science* 52, no. 2 (June 1989), pp. 288-302.

³²⁴ Millikan, "Biosemantics," pp. 281-283.

³²⁵ This is my formulation. I would like to capture, more or less, what Dretske understands by representation.

³²⁶ Millikan, "Biosemantics," pp. 284-285.

part of its proper function".³²⁷ Consider the heart for instance. The function of the heart is to pump blood, make a thumping sound, and when appropriately hooked to EKG, produce wiggly lines on the EKG display. But, only because of the selective effect of pumping blood do hearts exist. Therefore, the proper function of a heart is to pump blood.

(2) It is possible for a mechanism, trait or process to perform its proper function very rarely or even not to perform at all. It does not need to do what it was designed to do all of the time. The proper function of the tail of a sperm cell is to propel the cell to the ovum, yet a great many of them do not succeed. (3) It is possible for a mechanism, trait or process to have a proper function even though it does not perform it properly, i.e., if it is in a malfunctional state. We would not say that a heart of this person no longer has the proper function of pumping blood if that heart is malfunctioning.³²⁸ According to Millikan, the concept of proper function, the central concept in her version of teleosemantics,³²⁹ is a normative

³²⁷ Rowlands, *Teleosemantics*, p. 1.

³²⁸ All these three items come from Rowlands' interpretation. See Rowlands, *Teleosemantics*, p. 1. A defective heart here may not perform its proper function, but we still say that pumping blood is its proper function. On the other hand, a molecular duplicate of me that emerges out of cosmic accident, like in the case of Davidson's swampman, can have mechanisms that are identical with the mechanisms of a biological system, but, Millikan believes, the molecular duplicate would not have the right history, and therefore, its mechanisms (a heart, a kidney, an eye or a brain) would not have a proper function. If something does not have a proper function, then it is not a biological category. This is very crucial for Millikan, because she thinks that mentality, at bottom, is a biological phenomenon. See Ruth Millikan, *Language, Thought and Other Biological Categories: New foundations for realism* (Cambridge, Mass.; London, England: MIT Press, 1984), p. 93. Here, Millikan says: "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... For the categories "heart," "liver," "eye," "brain," and also "idea," "belief," and "intention," are proper function categories, defined in the end by reference to long-term and short-term evolutionary history, not present constitution or disposition." Italics are Millikan's. See also Barbara von Eckardt, *What is Cognitive Science?* (Cambridge, Mass.: London, England: MIT Press, 1996), pp. 223-224.

³²⁹ It is actually the central concept, although by different names, in all versions of teleosemantics.

concept that explains any occurrence of representational relations by virtue of what the mechanism in question related to representation consumption should do, not of what it does in general.

Then, after all this criticism of purely causal and informational theories, what is the ground of a representational relation for Millikan? Let's dwell more on the concepts in her toolbox such as proper function, normal condition and consumer mechanisms. Turning back to the definition of proper function, there are two essential points worth noticing about the way to interpret proper function of mechanisms. First of all, the first part (i) of Millikan's definition of proper function F indicates that F exists due to the selectively advantageous outcomes it manifested, no matter how rarely, for the survival and proliferation of the system/being in question. According to Millikan's point of view, this is a necessary condition for an item to have a *direct* proper function, and having a *direct* proper function is deeply connected to the evolutionary history of the relevant mechanism in question.³³⁰ *Derived* proper function (which is expressed by the second clause (ii) in the definition), on the other hand, is meant to express the function of any item that depends on the mechanism whose proper function is to produce items of that kind. When a chameleon is seated on a particular spot, then the mechanism responsible for arranging and distributing pigmentation on the skin of the chameleon allows the chameleon to fit with its immediate environment. We say that the mechanism has the proper function of arranging and distributing pigmentation. Also, we need to consider the

³³⁰ Millikan, "Biosemantics," p. 284; Rowlands, Teleosemantics, pp. 1-2.

particular state or item in the chameleon's skin; because it has a derived proper function such that it is dependent on the pigment arranger mechanism whose proper function is to produce that particular state or item in the chameleon's skin. This is important because it captures the fact that a *direct* proper function F is always related to its earlier patterns, its ancestral effects, whereas it is possible for a *derived* proper function F to have a first-generation instance; i.e., it does not depend on its earlier patterns. For instance, consider a chameleon seated on a particularly unique environment where none of its predecessors has ever been placed. In this case, the skin color of chameleon will instantiate a particularly unique pattern, and the state of chameleon's skin therefore will have the first-generation instance of a derived proper function.³³¹

In passing, we must keep in mind that direct proper function of a mechanism need not to be understood only in terms of the genetic selectional history of that mechanism, despite the fact that genetic selection is the paradigm case for direct proper functions. It is possible for some beings/systems to have feedback mechanisms through which those systems are able to learn from previous experiences and memory storage. We can easily say for those kinds of beings/systems that they are designed to learn through experience and modify their behavior in accordance with environmental data and pressures. Hence, design or proper function should not be understood only in terms of innateness of some item.³³²

³³¹ See Justine Kingsbury, "A Proper Understanding of Millikan," *Acta Analytica* 21, no. 3 (2006), pp. 23-40; Rowlands, *Teleosemantics*, pp. 1-2.

³³² Millikan, "Biosemantics," p. 284.

Millikan's notion of a normative description and normal condition is different from other senses of the term "normal" to the extent that her notion is all about the historically, perhaps very rare, performance of a particular function "when it was properly performed".³³³ In other words, the term "normal" does not have any connection with "what is typical or average or even, in many cases, at all common".³³⁴ So, for instance, by a normal condition for performance of a function, Millikan understands the presence of conditions where a particular function was performed properly.³³⁵ Normal explanation or description then is meant to give an account for the reason behind something's acquiring a particular function throughout its evolutionary history and normal condition for performance of a function must be included in that explanation.³³⁶ For Millikan, it is important to read "normal condition" and "normal explanation" as related to the selectional history of organisms and beings, and not related to the statistical average over the numerous performances of any given function of mechanisms of those organisms and beings. First of all, according to Millikan, "many functions are performed only rarely".³³⁷ To use Millikan's example, only few caterpillars manage not to be eaten even though they have protective coloration that enables them not to be seen by a predator. Secondly, "many proper functions only need

³³³ Ibid., p. 284.

³³⁴ Ibid., p. 285.

³³⁵ Ibid., p. 285.

³³⁶ Ibid., p. 285.

³³⁷ Ibid., p. 285.

to be performed under rare conditions".³³⁸ Millikan gives the example of the vomiting reflex. The mechanism whose proper function is to trigger vomiting exists because of selectively advantageous outcomes it had on the (very rare) occasions when a body was poisoned. Only under the very rare condition such as in the case of the poisoning must the vomiting reflex be performed.³³⁹ If the vomiting reflex were to be performed frequently, it would completely destroy the function of a nutritive system and thereby lead to the decay and inevitable death of organisms.³⁴⁰

After having outlined Millikan's terms such as "proper function" and "normal condition", now I shall present Millikan's argument about the role of consumer mechanisms. Millikan actually agrees with Dretske and Stampe in two main points: 1) Appeal to teleology, function is important for the project of naturalizing the intentional content of mental states, 2) What gives something a representational status is "that its function is to represent".³⁴¹ But, Millikan does not believe that it is a good and fruitful method to consider these insights in line with representation production. We must focus on the story about representation consumption. According to Dretske's account, when a honeybee performs its dance, then she produces a representation about the distance and location of nectar. For Millikan, however, this complex dance performance of a honeybee does not mean or represent anything at all

³³⁸ Ibid., p. 285.

³³⁹ Ibid., p. 285.

³⁴⁰ Millikan does not explicitly say this, but I believe this is actually what she wanted to say at the end.

³⁴¹ Millikan, "Biosemantics," pp. 283-284.

unless there is another honeybee or a group of honeybees, who can interpret and consume that representation in that honeybee's ecosystem such that, if that representation was in accord with reality (under "normal" conditions), it enables that system to survive and promote their generation. For Millikan, for anything to function as a representation, it must be used or consumed by the system in which it is embodied and produced.³⁴² To understand Millikan's point, let us then focus on the following question: What is it for a system to use a representation as a representation?³⁴³

If it is actually one of a system's functions to produce representations, as we have said, these representations must function as representations for the system itself. Let us view the system, then, as divided into two parts or two aspects, one of which produces representations for the other to consume. What we need to look at is the consumer part, at what it is to use a thing *as* a representation. Indeed, a good look at the consumer part of the system ought to be all that is needed to determine not only representational status but representational content. We argue this as follows. First, the part of the system which consumes representations must understand the representations proffered to it. Suppose, for example, that there were abundant "natural information" (in Dretske's sense) contained in numerous natural signs all present in a certain state of a system. This information could still not serve the system *as* information, unless the signs were understood by the system, and, furthermore, understood as bearers of whatever specific information they, in fact, do bear. (Contrast Fodor's notion that something could function like a representation without functioning like a representation of anything in particular.) So there must be something about the consumer that *constitutes* its taking the signs to indicate, say, *p*, *q*, and *r* rather than *s*, *t*, and *u*. But, if we know what constitutes the consumer's *taking* a sign to indicate *p*, what *q*, what *r*, etc., then, granted that the consumer's takings are in some way systematically derived from the structures of the signs so taken, we can construct a semantics for the consumer's language. Anything the signs may indicate qua natural signs or natural information carriers then drops out as entirely

³⁴² Ibid., pp. 283-284.

³⁴³ Ibid., pp. 285-286.

irrelevant; the representation-producing side of the system had better pay undivided attention to the language of its consumer. The sign producer's function will be to produce signs that are true *as the consumer reads the language*.³⁴⁴

This is one of the passages in “Biosemantics”, the article that is found, for many people, to be like the précis of Millikan’s much more complex and extensive account in *Language, Thought and Other Biological Categories*.³⁴⁵ But, the paragraph above from “Biosemantics” is really very hard to digest. Fortunately, Neander³⁴⁶ offers her interpretation of Millikan’s general argument about the necessity of consumer mechanisms. For anything to become a representation, it must be used as a representation by the consumer mechanism(s) of the system where it is produced. Otherwise, we need to accept that there could be representational states of systems without being used as a representation of anything in particular. It is absurd for something to be a representation without it representing anything.³⁴⁷ For this reason, it is necessary that representation is used as a representation by the consumer mechanisms.³⁴⁸

³⁴⁴ Ibid., pp. 285-286. Italics belong to Millikan.

³⁴⁵ Ruth Millikan, *Language, Thought and Other Biological Categories*: New foundations for realism (Cambridge, Mass.; London, England: MIT Press, 1984).

³⁴⁶ Neander, Karen. Spring 2012. Teleological Theories of Mental Content. Available [online]: “<http://plato.stanford.edu/archives/spr2012/entries/content-teleological/> [Spring 2012]”.

³⁴⁷ Ibid.

³⁴⁸ For this point, also see Ruth Millikan. “Biosemantics.” In *The Oxford Handbook of Philosophy of Mind*, edited by Brian McLaughlin, Ansgar Beckermann and Sven Walter (Oxford: Clarendon Press; New York: Oxford University Press, 2009). Millikan uses the same title “Biosemantics” both in her famous article in *Journal of Philosophy* and *The Oxford Handbook of Philosophy of Mind*. So, unless noted otherwise in the rest of this

From the long quotation above, there are two initial assumptions to be noticed: 1) For anything to be a representation, it must occur in a system. In other words, there is no such thing as a representation independent of any system (of representation) at all; 2) For anything to be a representation, there must be a consumer mechanism of the system, in which that representation is produced, and this consumer mechanism must read and understand that thing as a representation about a particular thing/state of affairs and modify behavioral options of the system in accordance with the environmental pressures and conditions outside the system. It seems to me that Millikan rests her teleosemantic account on these two assumptions; but still we do not have a clear answer for the following question: what does exactly determine the content of representation?

According to Millikan's teleosemantics, "the content of a representation is linked to the performance of the proper functions of its consumers".³⁴⁹ Consider again Millikan's example of a honeybee dance: The performance of a honeybee dance by a single honeybee does not mean anything unless that performance is interpreted by watcher/interpreter bees which constitute a consumer mechanism of the system of communication for members of that honeybee group. So, in this case, the bee performing this dance is producing a representation and the other bee(s) watching and interpreting that performance is

chapter, I will refer to Millikan's article titled "Biosemantics" in *Journal of Philosophy*, not the one in *The Oxford Handbook of Philosophy of Mind*.

³⁴⁹ Neander, Karen. Spring 2012. Teleological Theories of Mental Content. Available [online]: "<http://plato.stanford.edu/archives/spr2012/entries/content-teleological/> [Spring 2012]".

consuming a representation provided that that honeybee dance, on that occasion, gives a true representation of the location and distance of nectar; because only with a correspondence between representation and its represented it is possible for consumer mechanisms to function properly. Of course, this is a very wide perspective about a system whose consumer and producer mechanisms are instantiated by different beings. It is also possible to think of consumer and producer mechanisms as two aspects of the same system/being rather than as two different individuals/groups. In any event, it goes by definition that “a consumer is a system that normally exploits the mapping between a representation and its represented in the performance of its proper function, where the normality is teleological and not statistical”.³⁵⁰

For Millikan, content of representation is not directly determined by proper function of consumer mechanisms. This is one of the difficult points at Millikan’s thought. It is true that content is connected to the proper function of a consumer mechanism, yet this, alone, does not determine content:

Note that the proposal is not that the content of the representation rests on the function of the representation or of the consumer, on what these do. The idea is not that there is such a thing as behaving like a representation of X or as being treated like a representation of X. The content hangs only on there being a certain condition that would be normal for performance of the consumer's functions-namely, that a certain correspondence relation hold between sign and world-whatever those functions may happen to be. For example, suppose the semantic rules for my belief representations are determined by the fact that belief tokens in me will aid the

³⁵⁰ Neander, Karen. Spring 2012. Teleological Theories of Mental Content. Available [online]: “<http://plato.stanford.edu/archives/spr2012/entries/content-teleological/> [Spring 2012]”.

devices that use them to perform certain of their tasks in accordance with a normal explanation for success only under the condition that the forms or "shapes" of these belief tokens correspond, in accordance with said rules, to conditions in the world. Just what these user tasks are need not be mentioned.³⁵¹

The idea here seems to be the following: A consumer mechanism has a proper function; because previous tokenings of the type brought about selectively advantageous effects that help the system to survive, reproduce and fit in its natural habitat. For instance, ancestral frogs had ancestral digestive systems; and it is because those ancestral digestive systems contribute to the preservation and proliferation of frog population that frogs now have digestive systems. In any event, the proper function of a consumer mechanism is to enable the organism to survive, fit in its environment and promote its species for future generations, and it can normally realize this function, perhaps on rare occasions, only if there is a correspondence between representation (consumed by the system) and its represented. Then, content of representation is determined by a normal condition in which any consumer mechanism is able to function properly. In sum, in order to determine content of representation, there must be a normal condition where there is correspondence / mapping between the representation and the world such that a consumer mechanism is able to function properly.³⁵²

Despite the fact that it is particularly difficult to understand Millikan's emphasis on the idea of benefit-based approach, her theory

³⁵¹ Millikan, "Biosemantics," p. 287.

³⁵² Ibid., pp. 285-288. See also Neander, Karen. Spring 2012. Teleological Theories of Mental Content. Available [online]: "<http://plato.stanford.edu/archives/spr2012/entries/content-teleological/> [Spring 2012]".

gets the upper hand, in contrast to Dretske's informational theory, about the way to deal with cases of misrepresentation, those cases where the represented object does not exist or is not the way as it is represented to be. On Dretske's account, reliable correlation is what constitutes the relation of indication/representation between the source and the receiver. According to Millikan, this is not required for determining the content of representation. In Millikan's version of the teleological theory, the content of representation is determined over the selection of the proper function of the consumer mechanism under normal conditions. A representation, R, can still represent (what it is supposed to represent) the environmental feature of the world, E, even though there was never reliable correlation between R and E. In other words, it is sufficient that there has been enough correlation between tokenings of R and E under normal conditions such that the consumer mechanism, by making use of this representation, helps the system survive, reproduce and fit in the environment and consequently the proper function of the consumer mechanism has been selected. Furthermore, Dretske needed to make a distinction between training period and post-training period in order to explain the possibility of misrepresentation in his account.³⁵³ However, Millikan's theory does not have the same requirement to make a distinction as such; because it allows that misrepresentation can occur both in the training and post-training periods.³⁵⁴ In a nutshell, Millikan

³⁵³ I have remarked on the difficulties with this strategy in the previous chapter.

³⁵⁴ Ibid., pp. 288-290. See also Neander, Karen. Spring 2012. Teleological Theories of Mental Content. Available [online]: "<http://plato.stanford.edu/archives/spr2012/entries/content-teleological/> [Spring 2012]".

seems to say the following: once the representation mechanism has been selected to represent a certain situation, the content of representation is automatically fixed. So, for instance, if the proper function of that mechanism, M, is to represent a danger, say an approaching predator, and if M is triggered when there is no real threat outside, then it is simply malfunctioning. So, when M malfunctions, it misrepresents a certain situation.

It is often asserted that the teleological notion of function cannot completely fix the content of a state, so it leads to the problem of indeterminacy of content.³⁵⁵ Let us illustrate this criticism with Dretske's example of marine bacteria: Consider marine bacteria in the northern hemisphere, which have something called magnetosomes that function as the navigator for that organism in order to locate and move towards the magnetic north pole, and this eventually helps that bacteria to promote its existence; because the magnetic north pole is equivalent to oxygen-free water for that northern hemisphere bacteria—in southern hemisphere marine bacteria, these magnetosomes have the functional role to navigate them toward geomagnetic south, so southern hemisphere bacteria have their magnetosomes reversed. Swimming in the oxygen-rich surfaces of water means death for the bacteria. Now, it first appears that the function of these magnetosomes is to show the direction of oxygen-free water. So, we say that M has a representation OXYGEN-FREE WATER, THIS WAY.³⁵⁶ But, one can easily fool this mechanism by putting a magnet bar

³⁵⁵ This is also called "the problem of disjunctive content".

³⁵⁶ In the literature, when people refer to the content of representation, they generally capitalize the letters. So, I am following this tradition here.

nearby.³⁵⁷ Then, it turns out that the real function of this mechanism is not actually relevant with the direction of oxygen-free water; because when a magnet bar is placed nearby, then it will only respond to magnetic field and it may bring the organism to an oxygen-rich level of water and cause its death. So, we say this time that M has a representation MAGNETIC FIELD, THIS WAY.³⁵⁸ For Dretske, there is an important problem here, because there are two ways of reading the function of magnetosomes. On the local reading, we are supposed to say that the function of magnetosomes is to represent the direction of the surrounding magnetic field; but on the global reading, we are supposed to hold that the function of magnetosomes is to represent the direction of oxygen-free water. In other words, it is not really clear how to attach a univocal representational content to the functions of mechanisms producing representations. On the local reading, if you take the case of hemispherically transplanted bacteria into account, and since their magnetosomes still respond to magnetic field, then we are supposed to say that there is no case of misrepresentation here. But, on the global reading, since we attribute to magnetosomes the function to represent the direction of oxygen-free water, then hemispherically transplanted bacteria will be the victims of misrepresentation.³⁵⁹

³⁵⁷ Or alternatively, if you transplant southern marine bacteria to northern hemisphere, then it will cause the death for those bacteria.

³⁵⁸ The same result occurs for hemispherically transplanted bacteria even though their magnetosomes still respond to magnetic field in its surroundings.

³⁵⁹ Robert Cummins, *Meaning and Mental Representation* (Cambridge, Mass.: MIT Press, 1989), pp. 72-73; Millikan, "Biosemantics," pp. 290-291; Fred I. Dretske, "Misrepresentation." In *Mental Representation: A Reader*, edited by Stephen P. Stich and Ted Warfield (Oxford, UK; Cambridge, USA: Blackwell, 1994), p. 164; Rowlands, *Teleosemantics*, pp. 3-5; Kingsbury, "A Proper Understanding of Millikan," pp. 23-40. Robert Cummins explains the situation by the terms "conservative" and "liberal"

According to Millikan, the teleological theory of representation is good at these kinds of cases. Millikan says that it is not good to look at the production story; we need to look at the consumption story of representation. In other words, it is important to understand what the consumer mechanism is designed to do. In the case of these marine bacteria, the proper function of the magnetosomes is not selected due to its navigation effect of directing towards the magnetic north pole, or simply any magnetic field. The proper function of the magnetosomes is selected because of its navigation effect of directing towards the oxygen-free water and only in this normal condition where there was a mapping relation between magnetosomes and the environment was the bacteria able to survive. Therefore, the content of representation is not MAGNETIC FIELD, THIS WAY, but OXYGEN-FREE WATER, THIS WAY. This is another way of saying that Millikan believes that her account is a success in the attempt to provide univocal representational contents.³⁶⁰ In Chapter 7, I will revisit Millikan's account and talk about the controversy whether her account can solve the disjunction problem.

It appears, in this discussion of Millikan's teleosemantic views, that there is something radically wrong with the basic causal and informational theories of content, because they fall short of giving an account for the question "What is the nature of mental representation?" Millikan believes that her theory not only gives good clues about the nature of mental representation, it also successfully handles the general

readings. I prefer to explain the situation through the terms "global" and "local" interpretations/readings. This is only a matter of preference, not any slightest difference comes out.

³⁶⁰ Millikan, "Biosemantics," pp. 290-291.

problems of naturalist theories of content such as misrepresentation and indeterminacy. For these reasons, she believes that her account is successful in shedding light on the nature of mental representations.

CHAPTER VII

NATURE OF MENTAL REPRESENTATION

“What a strange machine man is!” he said, with astonishment. “You fill him with bread, wine, fish, radishes, and out of him come sighs, laughter and dreams. Like a factory. I am sure there is a sort of talking-film cinema in our heads.”³⁶¹

The purpose of this chapter is divided into three sections. First, I will begin with the initial formulation of the task concerning the definition of representation. As far as I know, there are mainly three options in the history of philosophy, which are proposed to give a definition of representation. In the first section of this chapter, I will present and show the troubles with the first two options. In the second section, I will remark on the trouble with the contemporary literature on mental representation and I will point out the trouble with teleosemantics within this context. In the last section, I will focus on the third option, the one that Aristotle offers in *De Anima*. I believe that understanding the soul of *De Anima* is a good direction towards understanding the very nature of mind and representation, something that both functionalist and post-functionalist positions fails to reach.

³⁶¹ Nikos Kazantzakis. *Zorba the Greek*, translated by Carl Wildman (London: Faber and Faber, 1961), p. 258.

The Task and Theories

As I said before, the question about the nature of mental representation / content³⁶² is the following one: what is the ground of mental representations? In virtue of what can an inner state of a subject (of mind) represent and think about the world? Then, let's suppose that the task is the following one:

S is a subject (of mental states) and has a state, X, that represents Y (object, event, relation, situation, fact outside S in his environment) iff _____

Let's take the formulation below as shorthand for the above:

(R) X represents Y iff _____

The task here is to fill in the blank space. Of course, there are two important demands. The first one is to tell something informative, namely, the account must be a non-circular one. So, for instance, the following is non-informative and useless: X represents Y iff there is a representation relation between X and Y. Saying that X represents Y because of a representation relation between X and Y is actually equal to saying nothing informative about the ground of representation. The second demand is to give a naturalistic account of representation, one that describes representation in terms of naturalistic facts (e.g., physical, chemical or biological facts).³⁶³ In other words, during the explanation of

³⁶² For an ingenious presentation of contemporary theories of content / mental representation in a vivacious literary and dialogue form, see Deborah Brown, "A Furry Tale About Mental Representation," *The Philosophical Quarterly* (October 1996), pp. 448-466. In the literature, the terms "theory of content" and "theory of mental representation" are often used interchangeably. For this point, see Crane, *The Mechanical Mind*, p. 25.

³⁶³ Crane, *The Mechanical Mind*, p. 174. The demand for naturalistic explanation has been one of the touchstones of contemporary philosophy of mind.

representation we should not appeal to non-natural elements such as intention, convention and so forth. The account must present “naturalistic” necessary and sufficient conditions for representation. According to Fodor, there are mainly two options in the history of philosophy that are said to be satisfying those two demands. These two main options, for Fodor, are “the resemblance account” and “the causal account”.³⁶⁴ Contrary to Fodor, I believe that we have a third option as well. There are, so far as I could see, mainly three kinds of options:

(R1) X represents Y iff there is (some sort of) resemblance between X and Y.

(R2) X represents Y iff there is a causal relationship between X and Y.

(R3) X represents Y iff (i) there is a causal relationship between X and Y, and (ii) there is (some sort of) a resemblance relationship between X and Y.

Let’s go over them one by one, and see the issues with each. There are too many troubles with R1. I will not go over each one of them.³⁶⁵ It is enough to deal with the most important issue here. No matter how you interpret R1 (either you take resemblance as literal or non-literal resemblance), you will end up with something really strange, that is, causation is not part of the story about mental representations. One may say that mental representations are, at bottom, the products of one’s

³⁶⁴ Jerry Fodor, “Semantics, Wisconsin Style.” In *A Theory of Content and Other Essays* (Cambridge, Mass.: MIT Press, c1992), p. 33. See also Crane, *The Mechanical Mind*, p. 175.

³⁶⁵ For an extensive discussion of resemblance theories of mental representation, see Crane, *The Mechanical Mind*, pp. 13-20; Robert Cummins, *Meaning and Mental Representation*, pp. 27-34; Jerry Fodor, “Semantics, Wisconsin Style,” pp. 32-34; Hilary Putnam, *Reason, Truth and History*, Chapter 1; Kim Sterelny, *The Representational Theory of Mind: An Introduction* (Oxford, Mass.: Blackwell, 1990), pp. 112-113.

causal interaction with the world. That is to say, mental representations³⁶⁶ ultimately rest on the previous causal connections with the world.³⁶⁷ It is really very hard to resist such a common-sensical intuition. Furthermore, if you do not include causation into the story about mental representations, then it is really very difficult to understand how R1 can give an account of the singularity of a representation. For instance, we say that when a subject represents a red rose in his vicinity, he must be, first of all, representing a particular red rose in his vicinity that is causally responsible for the production of the relevant mental representation in his mind. So, there is a difference between representing “this rose” and representing “rose” in general. “This rose” represents only this one. But, if one holds R1 to be true, then he is forced to admit that when a subject represents a red rose, he must be representing not any particular red rose in his environment. Instead, he must be representing all the resembling actual and possible red roses (to which he is supposed to have no causal connection). So, it does seem that R1 will not be able to explain the singularity of a representation in perception.³⁶⁸ For this reason, it seems that part of the story about mental representations must ultimately rest on the causal connection between subject and world. So, let’s move on to R2.

³⁶⁶ As a matter of simplicity, people often refer to mental representations by the capitalized words. So, I will follow this tradition in the literature. Consider the following example. The mental representations GOLD and MOUNTAIN may be the products of the direct causal interaction with the world, but GOLDEN MOUNTAIN is not the result of direct causal interaction with the world. It is a kind of mental representation that can be created in the faculty of imagination. So, as a Humean strategy, one can say that mental representation GOLDEN MOUNTAIN ultimately rests on the previous mental representations of gold and mountain, which are presumably the immediate products of the causal interaction between mind and world.

³⁶⁷ Crane, *The Mechanical Mind*, pp. 175-176.

³⁶⁸ For this point, see Jerry Fodor, “Semantics, Wisconsin Style,” p. 33.

In general, most contemporary philosophers think that truth about the nature of mental representation lies somewhere around R2.

“[R]epresentation is ultimately a causal relation or, more precisely, representation is based on certain causal relations.”³⁶⁹ Let’s then begin our analysis of causal theories of mental representation. It is useful to begin with the simplest form of the causal theory, what Fodor dubbed as “the crude causal theory” (henceforth, shortly as CCT).³⁷⁰ CCT makes the following claim:

(CCT) X represents Y, when and only when Y is the cause of X.

Issues with CCT are actually very obvious. First, it is possible for a subject to have mental representations about Y and yet it may not be even true for each of these mental representations that they occurred as a result of seeing Y. Consider the following example. Lately, I have sleeping issues, so I found a method to ease and relax my mind and, eventually, help me to get sleep. I count sheep before going to sleep. I imagine them jumping on the fence of my garden, and I count each one of them who succeeds in getting to the other side of the fence. But, there is no sheep in my bedroom. So, there is no causal connection between a sheep and my representation of sheep.

Secondly, it does not have to be true at all that when I am affected by Y, my mental state needs to represent Y. Consider the following example. When I was a schoolboy, I was walking down the hill near my house in Cyprus. Notice that it was the dark time of the day, and I saw something and got really very afraid. My mental state of fear in this case

³⁶⁹ Crane, *The Mechanical Mind*, p. 175.

³⁷⁰ Jerry Fodor, “Meaning and the World Order.” In *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, Mass.: MIT Press, 1987), pp. 99-102.

might occur as seeing a sheep in a dark night. But, I was not afraid because of seeing a sheep. There is nothing to be scared of seeing a sheep. A sheep, from the perspective of a child, is actually a very adorable animal. The truth was that I represented it as a wolf, and that was actually the real reason for the occurrence of my mental state of fear. So, a sheep may cause my mental state of fear, but what I represent may not be the same as the cause of my mental state. Instead of representing a sheep, which is the cause of my mental state, I have a mental representation of a wolf. So, it is not true that when Y is the cause of X, X needs to represent its cause, nor is it true that when X represents Y, there must be causation involved between X and Y.³⁷¹

Of course, these counter-examples are meant to show the weak spots of CCT. But, this should not be taken to imply, for many people, that any kind of causal theory couldn't succeed in shedding some light on the nature of mental representations. As a matter of fact, understanding the source of the problem can give some important insights as to develop and take CCT to a next level. In this respect, the theories of content/mental representation from the 1980s to the 1990s can be regarded as the attempts to find a better causal theory than CCT.³⁷² In the

³⁷¹ These two examples come from Crane, *The Mechanical Mind*, p. 176. I have made some modifications in Crane's original example. Similar counter-examples for CCT can also be found in Jerry Fodor, "Meaning and the World Order," pp. 101-102; Robert Cummins, *Meaning and Mental Representation*, pp. 35-75; Chris Eliasmith, "An Introduction to Contemporary Theories of Content" [Unpublished]. Available [online]: "http://www.arts.uwaterloo.ca/~celiasmi/courses/old_courses/Fall01/Phil600Repn/papers/eliasmith.2001.pdf"

³⁷² So, Dretske's indication theory, Papineau and Millikan's versions of teleosemantic theory, Fodor's asymmetric dependence theory can all be interpreted as the refined causal theories of content.

preceding chapters, I presented two of those attempts. Let's revisit them and see how they deal with the issues of CCT.

Inspired by Grice's idea of "natural meaning",³⁷³ Dretske believed that, instead of mere causal correlation, we need to explain mental representations in terms of natural regularities often found in nature.³⁷⁴ There is, for instance, always a natural regularity between smoke and fire. That's exactly the reason, for Dretske, why smoke is a natural sign/representation of fire. Dretske calls such relationships "reliable indication relationships", and he believes that mental representations, as presumably a species of natural representations, must be based on reliable indication relationships between representing vehicles (in the mind) and represented objects/events in the environment. So, the task for Dretske is to "explain the representational powers" of "the basic thoughts about the perceived environment, the basic drives (for food, drink, sex, warmth, etc.)" "in terms of a notion like indication"³⁷⁵ and, if there is some success achieved, to deal with complex cases later (like the thoughts about non-existent things, mathematical entities, future events, and so forth).³⁷⁶ So, Dretske, roughly, offers the following general theory of representation:

(Dretske1) X represents Y iff X reliably indicates Y.

³⁷³ See Paul Grice, "Meaning," *Philosophical Review* 66, no. 3 (Jul., 1957), pp. 377-388.

³⁷⁴ For the details about Dretske's theory, see Chapter 5. Also see Crane, *The Mechanical Mind*, p. 177.

³⁷⁵ Crane, *The Mechanical Mind*, p. 177.

³⁷⁶ This has been one of the typical background assumptions of the causal theories in the late twentieth century philosophy of mind. Explanation of complex thoughts must come later. I will come back to this point in the next section. See Crane, *The Mechanical Mind*, pp. 177-178; Hilary Putnam, "Does Evolution Explain Representation?" In *Renewing Philosophy* (Cambridge, London: Harvard University Press, 1992), pp. 22-24.

There are two big issues with Dretske's approach. Firstly, according to this view, mental representations are supposed to be reliable indicators of things, but it is really very difficult to see how this approach can ever explain the cases of misrepresentation. Let's call this "the problem of misrepresentation". Secondly, one may argue for the claim that "there are many phenomena which are reliably causally correlated with mental representations, yet which are not in any sense the items represented by them".³⁷⁷ Let's call this "the disjunction problem" for the reason that will become clear later. Let's call the two problems together "the problem of error".³⁷⁸ I shall go over each one of them now.

Following Grice's idea, Dretske says that "the spots represent measles" due to the reliable indication relationship between spots and measles. "If there are no spots, then there is no measles".³⁷⁹ Let's think of the converse of this claim: Could there be spots without measles? Let's suppose that there could be spots on someone's face, very similar to the ones when someone has measles. Suppose that these similar spots are actually the consequence of some other disease, for instance smallpox. But, this time we need to say that those spots would be reliable indicators of smallpox. So, Dretske's approach is forced to admit that those similar spots would not be misrepresenting measles. For "they represent what they indicate, namely smallpox".³⁸⁰ It is obvious that if Dretske holds his approach to be true, then he is forced to admit that X represents only

³⁷⁷ Crane, *The Mechanical Mind*, p. 178. See also Jerry Fodor, "Meaning and the World Order," pp. 101-102.

³⁷⁸ Crane, *The Mechanical Mind*, p. 178.

³⁷⁹ *Ibid.*, p. 178.

³⁸⁰ *Ibid.*, p. 178.

whatever it reliably indicates, and it cannot ever “represent something it does not indicate”.³⁸¹ This is not good, because now we have a very important difference between the way natural representations (smoke representing fire, etc.,) represent and mental representations represent. In the case of natural representations, when X represents Y, both X and Y must be present. So, for instance, if smoke means / represents fire, then this guarantees the fact that both smoke and fire must be present. However, we actually do not have this sort of guarantee in the case of mental representations. “[F]ew mental representations guarantee the existence of what they represent”.³⁸² Then, it seems that Dretske’s view, at this stage, cannot ever give an account for the cases of misrepresentation.

In the literature, it is often thought that the problem of misrepresentation has a close cousin, the so called “the disjunction problem”.³⁸³ Suppose that I have a mental representation of a sheep when there is a sheep around me. Call these mental representations “S-representations”. S-representations are reliable indicators of sheep in my environment, and it is exactly because of this reason that they are S-representations. Also suppose that under certain conditions (for instance, at a distance, in bad light, etc.,) “I am unable to distinguish sheep from goats”.³⁸⁴ So, my inability to distinguish sheep from goats is very

³⁸¹ Ibid., p. 179.

³⁸² Ibid., p. 179. See also Robert Cummins and Pierre Poirier, “Representation and Indication.” In *Representation in Mind: New Approaches to Mental Representation*, edited by Hugh Clapin, Phillip Staines, Peter Slezak (Amsterdam; Boston: Elsevier, 2004), p. 24; Jerry Fodor, “Semantics, Wisconsin Style,” p. 34; Jerry Fodor, “Meaning and the World Order,” pp. 101-102.

³⁸³ Crane, *The Mechanical Mind*, p. 179.

³⁸⁴ Ibid., p. 179.

systematic. That is to say, S-representations are not only the effects of seeing a sheep around, but they are also the effects of seeing a goat (at a distance, in bad light, etc.). Then, according to Dretske's indication theory, it follows that, since reliable indication is the ground of mental representations, we have to admit that "both sheep and goats-in-certain-circumstances are reliably indicated by S-representations".³⁸⁵ Then, S-representations would automatically have disjunctive contents: "either a sheep is present or a goat-in-certain-circumstances is present".³⁸⁶ This, however, is not a welcome result. There is a major consensus in the literature that when I have a mental representation of something, the content must be unique and determinate, not disjunctive. Otherwise, if mental representations were to have disjunctive contents, then we would turn back to the same problem, namely, to give an account of misrepresentation. The ideal thing is to say that my S-representations must be misrepresenting a goat as a sheep. But, if we take Dretske's indication theory to be true, then, it seems, we will never be able to make room for misrepresentation. Disjunctive contents, as being reliable indicators of objects, will never provide the resources to distinguish the right representations from the wrong ones.³⁸⁷ Crane nicely sums up the whole issue:

The misrepresentation problem is that, if reliable indication is supposed to be a necessary condition of representation, then X cannot represent Y in the absence of Y... The disjunction problem is that, if reliable indication is supposed to be a

³⁸⁵ Ibid., p. 179.

³⁸⁶ Ibid., p. 179.

³⁸⁷ For an extensive discussion of the problem of misrepresentation and the disjunction problem, see Jerry Fodor, "Meaning and the World Order," pp. 97-127; Jerry Fodor, *A Theory of Content and Other Essays* (Cambridge, Mass.: MIT Press, c1992), pp. 31-87.

sufficient condition of representation, then whatever X indicates will be represented by X. If it is a sufficient condition for an S-representation to represent a sheep that it reliably indicates sheep, then it will also be a sufficient condition for an S-representation to represent a goat-in-certain-circumstances that it indicates a goat-in-certain circumstances. Whatever is indicated by a representation is represented by it: so the content of the S-representation will be sheep or goat-in-certain-circumstances.³⁸⁸

Obviously, the two problems are related. They are both aspects of the problem that, according to the indication theory, error is not really possible. The misrepresentation problem makes error impossible by ruling out the representation of some situation (measles) when the situation does not exist. The disjunction problem, however, makes error impossible by ruling in the representation of too many situations (sheep-or-goats). In both cases, the indication theory gives the wrong answer to the question 'What does this representation represent?'³⁸⁹

Then, how did Dretske ever deal with these issues? He believed that the answer lies somewhere in understanding how conventional representations ever misrepresent. Take the example of a map into consideration. For the sake of simplicity, let us say that there is a strict and steady convention among mapmakers such that blue ink always means/represents "water". So, for instance, if one puts blue ink on a certain region of a map and if there is no water in the place about which map is to give information, then we have a case of misrepresentation. Now, blue ink, according to the common convention, has acquired a particular information-carrying role. When it fails to perform that role, we will have a misrepresentation of certain facts about a certain location. Following this thread of thought, Dretske then claims that ontogenetic

³⁸⁸ Crane, *The Mechanical Mind*, pp. 180-181.

³⁸⁹ *Ibid.*, p. 181.

function acquisition and learning must be added to the causal story about mental representations.³⁹⁰ So, Dretske modifies his earlier view of general representation and proposes a solution for the problem of error:

(Dretske2) X represents Y iff X has the function to reliably indicate Y.

In order to explain how a person can get this function to reliably indicate something, Dretske then goes on to give an example about concept acquisition. The basic idea, as you will recall, is to distinguish the training from post-training period for learning concepts. In the training period, a student, say S, may be said to be receiving diverse sorts of stimuli, and with the help of his teacher he comes to make certain kinds of discriminations about them. After the training is over, Dretske believes, the student then gets the conceptual ability / function to distinguish, for instance, birds from airplanes. It is still possible, Dretske underlines, for S to use, in his box of concepts, a concept that is not triggered by the right cause, but only this time it will be regarded as wild tokening of that concept. Since S is, in the post-training period, a person who knows all about the concepts he has acquired, those tokenings accidentally triggering the concept are now counted as wild and will not sneak into the representational content of that concept. For instance, if the student in the post-training period mistakes a fox for a dog, then his concept "dog" will not mean "dog or fox"; it will still mean / represent dogs. In a nutshell, Dretske makes a distinction between what happens in the training period and what happens afterwards, and by means of this distinction, he claims that wild tokenings of a concept can only occur after the student knows everything there is about that concept. Those

³⁹⁰ I have previously discussed the details of Dretske's move here, so see Chapter 5.

wild tokenings are delicately differentiated from the normal tokenings of a concept. So, when a student, after the training period, utters a sentence "This is F", where the relevant object is G or the relevant object does not even exist, this will be an instance for him to misrepresent the situation. The problem of error, Dretske believes, can be solved in this way.

I have already dealt with Dretske's way of solving the problem of error.³⁹¹ There are many issues with his proposal. Firstly, it is really a very incredible assumption to say that learning concepts come to an end and we then acquire full authority on using concepts. Secondly, Dretske's example is in conflict with his big project of naturalization. As I think of Dretske's example more and more, I notice that the example implicitly appeals to many things such as the intentions of the teacher, going to school, use of the concepts in accordance with the conventions of the society, etc., things that violate the requirements of naturalism. A quick look at the example will reveal all these things to the surface. During the training period, a subject is trained to use "X" as a mental representation of birds. Dretske says that when this student graduates from "X" teaching school, and mistakes an airplane for a bird, then we would say that airplane is not one of the normal causes for "X". So, whenever any airplane triggers "X", then we would say that it is false tokening of "X" and we have a case of misrepresentation. But, it is possible to interpret the same situation counterfactually. If any airplane can cause "X" in the post-training period, then it could have been true of that airplane to cause "X" in the training period as well. Just imagine one day before the graduation the student looks at the sky; he sees an airplane and mistakes

³⁹¹ See Chapter 5 for the details.

it for a bird. So, instead of saying that the student learnt "X" means bird, we can equally say that the student might have learnt "X" means airplane or bird. Then, when the student in the post-training period utters the word "X" in the presence of airplane, we should say that there is no misrepresentation; because he had previously learnt that "X" means airplane or bird. In Dretske's example, however, you can feel that the teacher's intentions, conventions of the society, etc., play a central role in determining how and why the student learns that "X" means bird rather than "X" means bird or airplane. So, when you emphasize, in the example, the role of the teacher, schooling, conventions of the society about the meaning of a concept, you may get around the problem at the cost of failure of the project of naturalism, which is not a good trade. Naturalism is one of the big goals in Dretske's theory, so I do not think that he will ever consider abandoning it seriously. Last but not less, the learning story about the cases of misrepresentation can only tell us how misrepresentation can occur. Dretske's account straightforwardly implies that if a signal carries the information that s is F then s really is F. So, on this account, when my mental state represents that s is F, then it must be the case that s is really F. There is no way my mental state can misrepresent a certain situation. Dretske's theory already rules out misrepresentation, and it just does not help a bit to tell a story how misrepresentation can occur. Saying something about how misrepresentation can occur and solving the problem of misrepresentation in the theory are two different things.

One thing that Dretske diagnosed correctly was that mere nomic correlation/reliable indication is not enough to explain the ground of a

representing relationship between representing vehicles (in the mind) and represented objects/events in the environment. The very idea of “function”, at this point, seems to be the right route to get some progress. However, it seems that Dretske did not put much effort to elaborate the notion of function within the constraints of naturalism; this, in particular, becomes obvious with his example about concept acquisition. Instead of giving an account for how natural functions ever emerge he prefers to rely on the story about conventional functions and intentions of a teacher, and so forth, which determine the correct use of concepts. Millikan’s teleosemantic theory, in this regard, seems to be good at detecting the source of the problem and developing a notion of function that satisfies the requirement of naturalism, which is never to appeal to notions like intention, convention, and so forth during the explanation of mental representations. So, let’s now move on to Millikan’s theory.³⁹²

Millikan’s adaptationist approach, among the candidates so far, seems to be the best causal theory. Roughly, Millikan offers the following theory of mental representation:

(Millikan) X represents Y, iff X is a mechanism whose proper function (i.e., to represent Y) is determined by its evolutionary history such that it enables the subject, S, to survive, fit in its environment and reproduce.

As far as I understand Millikan, the key element of her theory is the concept of “proper function”. The proper function of biological mechanisms is determined by the evolutionary history of those organisms making use of the biological mechanisms. Consider the following example. A defective heart, for instance, may not perform its

³⁹² See Chapter 6 for the details.

proper function, but we still say that pumping blood is its proper function. On the other hand, a molecular duplicate of me that emerges out of cosmic accident, like in the case of Davidson's swampman, can have mechanisms that are identical with the mechanisms of a biological system, but, Millikan believes, the molecular duplicate would not have the right history, and therefore, its mechanisms (a heart, a kidney, an eye or a brain) would not have a proper function. This is very crucial for Millikan, because she thinks that mentality, at bottom, is a biological phenomenon. So, if a being does not fall under a biological category, like in the case of Swampman, then we cannot attribute thoughts to that being.

One of the strengths of Millikan's theory is that it finely locates the source of the problem in Dretske's indication approach. As I said before, Dretske believes that mental representations must be reliable indicators. However, Millikan argues for the claim that Dretske's indication approach gives us really a very wrong picture about mentality in general and mental representations in particular. Mental representations need not be thought as reliable indicators. In fact, the evolution proves that the living beings with fast-but-not-always-accurate representation mechanisms were much luckier than those which have slow-but-always-accurate representation mechanisms. It is exactly because of this reason that the former mechanisms have been selected. For Millikan, this is very important. Since we are biological beings, then we should understand first the proper function/teleological function of the mechanisms that help us survive, fit in the environment and pass our genes to the next generation. So, according to Millikan's view, the ground of mental

representations lies in the proper function of representation mechanisms whose function is determined by the evolutionary processes surrounding the biological beings. This, I believe, is the first step of Millikan's theory that locates the source of the problem in Dretske's indication approach.

Having provided a teleological notion of function, then Millikan goes on to explain the cases of misrepresentation. I believe that the other strength of Millikan's theory lies in its way of solving the problem of misrepresentation. Consider the following example. Let's say DOG stands for a mental representation of dogs. Typically, we want to say that DOG means/represents "dog". But, it is quite possible for a subject to have a mental representation DOG when, for instance, his perception is caused by a wolf-in-the-distance. So, how can teleosemantic theory give an account for the cases of misrepresentation? Millikan's response, as far as I understand, is the following one: First, there is no such thing as an isolated and useless mental representation DOG for the organism in which that mental representation has been produced. We must give an account of the teleological reason behind the emergence of mental representations. So, let's say that we are talking about a certain kind of species whose natural predators are dogs. Millikan says, this mental representation DOG here is determined over the selection of the proper function of the consumer mechanism under normal conditions. A representation, R, can still represent (what it is supposed to represent) the environmental feature of the world, E, even though there was never reliable correlation between R and E. In other words, it is sufficient that there has been enough correlation between tokenings of R and E under normal conditions such that the consumer mechanism, by making use of

this representation, helps the system survive, reproduce and fit in the environment and consequently the proper function of the consumer mechanism has been selected. So, once the representation mechanism has been selected to represent a certain situation, the content of representation is automatically fixed. For instance, if the proper function of that mechanism, M, is to produce and consume a mental representation DOG HERE, DANGER and if M is triggered when there is no real threat outside, then it is simply malfunctioning. So, when M malfunctions, it misrepresents a certain situation. This is, as far as I understand Millikan, her solution to the problem of misrepresentation. For her view, once the content is fixed, then the problem of misrepresentation evaporates. We can say that when the mechanism M is triggered in the presence of wolves, then it is simply malfunctioning. Since M has acquired the teleological/ proper function to represent dogs, it does not lose this function when there is no dog around. So, we would say that, when there is not any single dog around and M goes into a state DOG HERE, DANGER, then it is only malfunctioning. This is, for Millikan, the case for M to misrepresent a certain situation in the environment.

It does seem, after the discussion in this section, that Millikan's teleosemantic theory is the best candidate for explaining the ground of mental representations. In the next section, I will attempt to show the trouble in the contemporary literature, especially within the context of the question concerning the conceptual resources of teleosemantic theories.

The Trouble with the Frog

One of the working hypotheses behind the research strategies for probably most of the theories of mental representation in the late twentieth century was that any theory of mental representation must begin with the analysis of simple mental states/capacities (e.g., states that represent food, drink, danger and so forth), and if the theory in question achieves a considerable success in explaining those simple ones, deal with complex mental states/capacities later. As an example of complex mental states/capacities, people generally cite things like a meta-representational capacity (i.e., representation of representation),³⁹³ representation of mathematical objects (e.g., numbers, sets, functions and so forth),³⁹⁴ representation of non-existent objects (e.g., unicorns, Santa Claus and so forth),³⁹⁵ and representation of future events, and so forth.³⁹⁶

³⁹³ The role of a meta-representational capacity has been studied in connection with various kinds of mental phenomena such as intelligence, communication, consciousness and inferential mechanisms of intelligent beings. For a short introduction, see Dan Sperber, "Metarepresentation." In *The MIT Encyclopedia of Cognitive Sciences*, edited by Robert A. Wilson and Frank C. Keil (Cambridge Mass.; London, England: The MIT Press, 1999), pp. 541-543. For an extensive introduction to various themes about meta-representational skills, see Dan Sperber (Ed.). *Metarepresentations: A Multidisciplinary Perspective* (Oxford, New York: Oxford University Press, 2000). In general, the term "meta-representational capacity" is generally understood as a certain kind of capacity of a subject to recognize (either in his own or other's mind) that there are intentional/representational states such as beliefs and desires (Dan Sperber, "Metarepresentation," p. 541.). With the exception of primates, it is generally accepted that most animals simply lack meta-representational capacity, the capacity peculiar to highly intelligent social animals through which they can "interpret and predict the behavior of others by recognizing their mental states" (Dan Sperber, "Metarepresentation," p. 541.). For some people, we can find the instances of the most complicated kinds of meta-representational abilities in the social behaviors of human beings. For some people, this may even be taken to be the defining feature of consciousness. One thing is clear though: meta-representational capacity is definitely a highly complex intentional/representational trait (Dan Sperber, "Metarepresentation," p. 541.).

³⁹⁴ Crane, p. 177.

³⁹⁵ Crane, p. 177.

As an offspring of this working hypothesis, most of the debates in this era revolve around the questions about fixing the content of simple representational/intentional states. One problem, especially relevant with the conceptual resources of teleosemantics, is the following one:³⁹⁷ Let's take the tongue-snapping behavior of frogs as the effect of a simple representational state. As a standard teleosemantic explanation,³⁹⁸ one can say that the proper function of the representational mechanism in a frog's brain has been determined, under normal conditions (where flies are around), due to the fact that representing flies helped the frog to feed itself and, consequently, this helped promoting its existence and passing its genes to the next generation. However, Fodor and others³⁹⁹ believe that a teleological story about intentionality won't help to solve the disjunction problem. For Fodor, it is equally plausible to say that the frog snaps at anything that is an appropriately small and dark moving object in the frog's natural habitat. Then, we can say that the relevant mechanism actually represents FLY OR SMALL DARK MOVING

³⁹⁶ Crane, p. 177. For the discussion of this working assumption, also see Hilary Putnam, *Renewing Philosophy*, Chapter 2. In passing, I would like to make it clear that the list above for complex mental states/capacities are not meant to be complete, nor is it meant to give mutually exclusive types for complex mental states/capacities. For instance, having a meta-representational capacity may be regarded as intimately connected with the ability to represent things in the future. In a sense, while attributing intentional/representational states to other beings, we do not only try to understand their mentality, we also try to predict possible behaviors of other beings. For the relation between meta-representational capacity and representing future events, see Josef Perner, *Understanding the Representational Mind* (Cambridge, Mass.; London, England: MIT Press, 1991), p. 7.

³⁹⁷ See Jerry Fodor, "Theory of Content I." In *A Theory of Content and Other Essays*, pp. 64-82. The example belongs to Fodor.

³⁹⁸ Ruth Millikan, "Biosemantics," *Journal of Philosophy* 86, no. 6 (June 1989), pp. 281-297.

³⁹⁹ Hilary Putnam, "Does Evolution Explain Representation?" In *Renewing Philosophy* (Cambridge, London: Harvard University Press, 1992), pp. 19-34; Paul M. Pietroski, "Intentional and Teleological Error," *Pacific Philosophical Quarterly* 73, no. 3 (1992), pp. 267-81.

OBJECT. But then, it turns out that, contrary to the advertisement of the teleological story about mental representations, appealing to Darwin's theory of evolution does not really give us a univocal and determinate content about what is being represented.⁴⁰⁰ So, Fodor ends up with saying:

The Moral, to repeat, is that ... Darwin doesn't care how you describe the intentional objects of frog snaps. All that matters for selection is how many flies the frog manages to ingest in consequence of its snapping, and this number comes out exactly the same whether one describes the function of the snap-guidance mechanisms with respect to a world that is populated by flies that are, *de facto*, ambient black dots, or with respect to a world that is populated by ambient black dots that are, *de facto*, flies... *Darwin cares how many flies you eat, but not what description you eat under...* So it's no use looking to Darwin to get you out of the disjunction problem.⁴⁰¹

The problem about the indeterminacy of content/disjunctive content also seems to give rise to another difficulty. If, under one interpretation, we can choose to assign the representation FLY to the function of the frog's representation mechanism, and if, under another interpretation, we can plausibly choose to assign the representation SMALL DARK MOVING OBJECT, then it seems that what is being represented totally depends on one's preference of interpretation. Given another working hypothesis in the literature about the idea of naturalizing mind, this does not seem to be a good consequence at all. If the ideal goal is to naturalize the mind,

⁴⁰⁰ Fodor, "A Theory of Content, I." In *A Theory of Content and Other Essays*, pp. 71-72.

⁴⁰¹ *Ibid.*, pp. 72-73. Italics belong to Fodor.

then how can we be allowed to pick different interpretations for the same situation? This seems to jeopardize the project of naturalizing the mind.⁴⁰²

As far as I know, there have been very long debates about what really the frog represents when it feeds itself. On the one side, Millikan and others⁴⁰³ responded to the charge of indeterminacy / disjunction objection by pointing out the fact that the notion of *selection for* is a strictly causal notion and, contrary to the charges, it always makes it clear that the function of a trait is all about what that type of trait has been selected for. For instance, we can say that the function of a heart has been selected for circulating blood, not for its making heart beats. So, the function of a heart is all about circulating blood, not other things like making thumping sounds. So, Millikan says, the same thing must be said about representation mechanisms. The proper function of the representation mechanism of frogs is to represent flies, not any other similar things in the natural habitat of frogs. On the other side, some other people, including Fodor,⁴⁰⁴ have still insisted on the claim that even this response does not provide a fine-grained content such that it would help blocking

⁴⁰² Neander, Karen. Spring 2012. Teleological Theories of Mental Content. Available [online]: "<http://plato.stanford.edu/archives/spr2012/entries/content-teleological/> [Spring 2012]".

⁴⁰³ See Ruth Millikan, "Biosemantics," *Journal of Philosophy* 86, no. 6 (June 1989), pp. 281-297; Ruth Millikan, "Truth Rules, Hoverflies and the Kripke-Wittgenstein Paradox," *Philosophical Review* 99 (1990), pp. 232-53; reprinted later in Ruth Millikan, *White Queen Psychology and Other Essays for Alice* (Cambridge, MA: MIT Press, 1993), pp. 211-239; Ruth Millikan, "Speaking Up for Darwin." In *Meaning in Mind: Fodor and His Critics*, edited by Loewer, B. & Rey, G. (Cambridge, MA: Blackwell, 1991), pp. 151-165; Kim Sterelny, *The Representational Theory of Mind: An Introduction*, Cambridge, MA: Blackwell, 1990.

⁴⁰⁴ See Paul E. Griffiths, "Functional Analysis and Proper Functions," *British Journal for the Philosophy of Science* 44, no. 3 (1993), pp. 409-422; Paul E. Griffiths & R. Goode, "The Misuse of Sober's Selection for/Selection of Distinction," *Biology and Philosophy* 10, no. 1 (1995), pp. 99-107; Karen Neander, "Malfunctioning and Misrepresenting," *Philosophical Studies* 79, no. 2 (1995), pp. 109-141; Jerry Fodor, "Deconstructing Dennett's Darwin," *Mind and Language* 11, no. 3 (1996), pp. 246-262.

the possibility for the frog to represent FLY OR SMALL DARK MOVING OBJECT. For one can still say that it might be because of representing SMALL DARK MOVING OBJECT and the frog's being lucky to not to eat something dangerous in its natural habitat that might have determined the function of representing mechanism. In a nutshell, the suspicion is about whether evolutionary theory can be used to explain how representational contents of organisms are determined. Since natural selection is extensional, then it will be not clear why a subject confronting environmental pressure E adapts itself to E rather than another environmental pressure/ situation F, which can be said to be reliably co-extensive with E. If both environmental features, E and F, must go hand in hand for the job-description of a consumer mechanism, then there will be no account to fix representational contents. Thus, no way to distinguish behaviors of different kinds. In any event, the frog's simple representation has been the central debate concerning the conceptual resources of teleosemantic theories.

After reading all these sorts of lively debates about simple mental states within the context of teleosemantic theories, it is really disappointing to see that the literature got stuck with the first step (explaining simple mental states/ capacities) and did not even manage to move to the next step (explaining complex mental states/ capacities, which are much more interesting and important). For instance, why is not there a debate about the case of "the behavior of a rich old man buying an expensive meal in a restaurant for a younger woman"?⁴⁰⁵ An archenemy of teleosemantic theory could say that cases like this are not just instances

⁴⁰⁵ Crane, *The Mechanical Mind*, p. 196. This is Crane's example.

of adaptation on the ground that adaptation can explain biological aspects of beings, not their highly complex social activities that have been fixed by the rules and conventions of society and so forth.⁴⁰⁶ A defender of teleosemantic theory, for instance, could still insist that, even though this is a complex social activity, in the sub-personal level some of the evolutionary mechanisms (e.g., the old man's biological code to promote his species and pass his genes to the next generation, being attracted by a young woman because youth is a good indicator of fertility, and so forth) can play an important role in determining the representational content at least in the sub-personal level, albeit the subject is not conscious of it. Millikan and others, at this point, need not deny the import of cultural, social and linguistic practices behind the complex mental states.

Of course, one may go further and find a more puzzling/difficult case in which no selective mechanism might play any role in determining what is being represented. For instance, I wonder what would be the teleosemantic explanation for a person thinking that the Arab Spring has been a quite influential movement in the Middle East. It is not hard to find examples of this sort. For these kinds of cases, it seems that it is, I would say, nearly impossible to find a teleosemantic explanation. I do not see, in particular, how entertaining this thought and representing the world in such a manner could have been interpreted as the effect of a teleological advantage it has brought with respect to my survival, fitness and reproduction (three important determining conditions, according to teleosemantic theories, behind the function of representation mechanisms). Consider the following example. Right now I look at the

⁴⁰⁶ Ibid., p. 196.

sky above and start thinking of Milky Way galaxy. So, I am in a certain mental state that picks out the object very far away from the person living in this world. Again, I do not see any possible teleological explanation for representing Milky Way galaxy. First of all, Milky Way galaxy has nothing to do with the selection processes in this world. Secondly, thinking of Milky Way galaxy would not bring any selective advantage for me over other organisms in this world that do not have even the slightest idea about Milky Way Galaxy.⁴⁰⁷ Therefore, it appears that teleosemantics just falls short of giving any sensible account for the cases of complex mental states/capacities. Of course, one can still hold the claim that Millikan's teleosemantic theory is, at least, good at explaining basic thoughts and drives an organism has about food, drink, warmth, and so on. But, this should not be enough for a theory of representation. What we need in this literature is to find a theory capable of explaining all sorts of mental states, not only basic thoughts and drives. Let's now move on to Aristotle's account.

The Aristotelian Way⁴⁰⁸

In the previous section, I began with the formulation of the task concerning the definition of representation. As I said before, there are mainly three options in the history of philosophy, which are proposed to give a definition of representation. In the previous section, I only talked

⁴⁰⁷ Both examples belong to me.

⁴⁰⁸ In preparing this section, I greatly benefited from Chryssi Sidiropoulou (forthcoming), "A New Understanding of Aristotle's 'Dualism' in the *De Anima*," *Philosophical Inquiry*. Furthermore, I owe special thanks to Stephen Voss and Chryssi Sidiropoulou for their suggestions and criticisms.

about two of them and explained the trouble with those options, be it simple versions or more refined options. In this section, I will start talking about the third option, the option that Aristotle proposed. Aristotle's account differs from the post-functionalist accounts defended by Dretske and Millikan, because, unlike Dretske and Millikan, it is not built on a Platonic metaphysics. Furthermore, there is a very crucial point in Aristotle, which, I believe, indicates exactly the missing perspective in the views of Dretske and Millikan. In this section, first I will summarize some of the main points in Aristotle's account of mind as a representing capacity, and then I will focus on his distinction between philosophical definition and physical/natural definition. I believe that understanding this difference helps us understand the good direction towards understanding the nature of mental representation, and show the source of the chronic problem within post-functionalist naturalist attempts to understand mind and representation. At the end, I will claim that we have at least a (relatively) good criterion, as opposed to the criteria given by Dretske and Millikan, in Aristotle's theory that tells us exactly what is being represented in mind and what is being misrepresented in mind.

First, I would like to begin with some general remarks on the way Aristotle understands mind as a representing capacity. Contrary to the Platonic metaphysics in both original functionalism and post-functionalism, we can see in *De Anima* that Aristotle, most of the time, understands soul as an embodied being.⁴⁰⁹ Mind is only one of the powers of an individual being with a soul. As a corollary of his general

⁴⁰⁹ I have previously explained this aspect of Aristotle's account at some length. See Chapter 2.

metaphysical claim about the unity of form/ function and matter/ structure, Aristotle holds the claim that form/ function and matter/ structure of mental beings are deeply connected, contrary to the Platonic/ Cartesian metaphysical background operative in both functionalism and post-functionalism.

Let's go through some of the passages in *De Anima* where Aristotle's understanding of mind becomes very explicit. Let's first remember Aristotle's treatment of sensation. In one passage, Aristotle talks about the importance of having a sense organ.⁴¹⁰ If sensation were nothing but the mere causal/ material process in a patient object, then we would not have any way to differentiate the case of air picking up a smell from the case of an animal really smelling something. After all, if the criterion were to be the mere presence of sensible form, then we would have to say that there is perception in both cases. Just in the middle of this discussion, Aristotle makes his first move: in order to talk about sensation, a patient object must have a sense organ. In the case of air, there is not any sense organ to sense anything, so, Aristotle says, air is affected by the sensible form only so as to serve a medium of sensation for a subject with a sense organ. After this passage, Aristotle also talks about the physical capacity of sense organs. If the sensible form of the sense object is above or below the range of the physical/ structural capacity of sense organs, then a subject may either lose (permanently or temporarily depending on the degree of the effect) his power of sensation or may not sense anything. There is always a limit for the things to be sensed; this should be read in parallel to his discussion of the difference

⁴¹⁰ Aristotle, *De Anima*, 424a17-424b20.

between air picking up a smell and a sense organ smelling something. There is no limit for air; since it is not a sense organ of anything (subject of mental states), it does not matter whether the sensible form affecting it is too strong or not.

For Aristotle, sensation is deeply related to the physical/structural capacity of sense organs. The form/function and matter/structure of sensation are related in such a way that what is being sensed is partly determined by the physical capacity of sense organs. So, Aristotle would say, if two subjects, say S1 and S2, differ in the physical capacities of their sense organs, they will necessarily differ in the way of what can be sensed/perceived. In Aristotelian functionalism, I see the following point: if X is a mental state (e.g., sensation is, after all, only one kind of mental state), then what X does in a functional network/architecture is partly determined and constrained by the physical/structural feature of X. The function/form of X is not independent of the material of X. Only with the union of function/form and structure/matter of X, can we talk about the very nature of X. One just cannot sense anything if there is no sense organ up to the task, and the function of sensation can be understood only with its relation to the body. This is a very unifying perspective, one that is not found in original and post-functionalism in the twentieth century philosophy of mind.

In the case of sensation, one can easily notice the way Aristotle connects the form/function of sensation with the matter/structure of sensation. But, when Aristotle talks about mind, we can start to notice that this is a very complicated matter for Aristotle. One thing is clear: Aristotle does not want to assign an organ for the mind, nor does he

think of mind as inseparable from body. But, even though mind (both patient mind and agent mind) does not have any organ, this should not mean that mind is a completely abstract and independent substance.⁴¹¹ That is definitely not what Aristotle intended to do. Rather, even in the case of mind, Aristotle still thinks that mind is continuous with sensation and imagination such that what is being represented in mind is directly influenced by sensible forms in sensation and phantasms in imagination. So, in a sense, we can still talk about the fact that what is thought is limited to and determined by whatever (sensation and imagination) feeds the thought. Since sensation is limited with respect to the physical capacity of sense organs and imagination always requires sensation, then mind must also be limited to think about the forms presented and then abstracted by sensation and imagination.⁴¹² Therefore, Aristotle simply says, you cannot think of anything if the material for thinking is not provided by sensible forms in sensation and phantasms in imagination. As the body is essential for sensation so is it also essential for thinking.

In a nutshell, Aristotle understands mind, most of the time, as something embodied/embedded. As a corollary of this general attitude, I think Aristotle would say that if mind is a representing capacity, then it follows that this capacity must also be explained with reference to the material/physical limitations of a bodily being. This is, I take it to be, the Aristotelian functionalist way of understanding mind as a representing capacity. In Aristotle's account, there is a balance between the

⁴¹¹ Aristotle, *De Anima*, 429a21-27. See also 429b0-6. See Chapter 2 for the details.

⁴¹² Imagination, contrary to what is understood nowadays, is not a sort of creative power in Aristotle's account. One cannot imagine something that does not have a sensory basis. I owe this point to Chryssi Sidiropoulou.

form/function and matter/structure of a bodily mental being. Unlike in the Platonic tone of original and post-functionalism, Aristotle would not say that the form/function is all there is to the nature of mind (so he is not a functionalist in the contemporary sense), nor would he say that the matter/structure is all there is to the nature of mind (so he is not a reductive materialist).⁴¹³ For Aristotle, it is the being (“human ousia”) composed partially by the form/function and matter/structure that together determine the nature of mind as a representing capacity. Thus, mind cannot be reduced to or explained only in terms of form/function or matter/structure of a mental bodily being. I shall explain this crucial point in more details.

I think that the best place to understand Aristotle’s unifying perspective comes in his discussion of the affections of the soul in *De Anima*.⁴¹⁴ This is the place where Aristotle begins with the question whether the affections of the soul are separable from the body or not, and then he starts considering some of the cases like anger, fear, joy, love, hate, etc. In all these cases, it seems that they are not separable from the physical material of animals. In defining anger, Aristotle says, a physicist and a philosopher eventually offer very different sorts of explanation.⁴¹⁵ A physicist only looks at what physically happens when somebody is angry, and therefore he gives a very physical/material explanation. Anger, for a physicist, is nothing but boiling of the blood or heat about the heart. As for a philosopher, we can see a radically different sort of

⁴¹³ For an introduction to reductive and non-reductive physicalism, see Jaegwon Kim, *Philosophy of Mind* (Boulder, Col.: Westview Press, 1996), pp. 211-240.

⁴¹⁴ Aristotle, *De Anima*, 403a3- 403b19.

⁴¹⁵ Aristotle, *De Anima*, 403a30- 403 b2.

explanation. Anger, for a philosopher, is a desire for retaliation. The interesting point about this discussion is that while the former points out the matter/structure of anger, the latter points out the form/function of anger. I think that the correct way to read this discussion goes through Aristotle's unifying perspective such that none of the definitions alone can be taken to be complete as regarding the nature of anger. This should tell us something about the way Aristotle understands mind and representation.

As for the case of representation, a physicist/naturalist functionalist philosopher looks at only what physically happens in the body/brain of a subject. So, for instance, Millikan actually gives us a very physicalistic/functionalist definition for representation: S has a state X whose proper function is to represent Y, because X is a biological event in S's brain that is determined, by its evolutionary history, to represent Y. Aristotle, on the other hand, would probably not be satisfied with this sort of physicalist story about representation. Millikan's account only tells us why/how a representation has occurred the way it did. However, following Aristotle, if we are after a philosophical explanation for the nature of representation, then we should also look for a conceptual/formal analysis for the ground of mental representation. This is a different story from a physical story, because the project here is not to explain the physical occurrence of representation. So, according to this task, unlike Dretske and Millikan, we are supposed to give a formal criterion for what is being represented and what is being misrepresented in the personal level, not a material criterion for what occurs in the sub-

personal level (e.g., neuro-chemical, biological processes, etc.,) when I represent something.

I believe that we should understand Aristotle's definition of representation in the way he explains anger. So, the definition must capture both the material and formal component of representation.

Aristotle roughly makes the following point:

(Aristotle) X represents Y iff (i) there is a causal relationship between X and Y (material/physical criterion), and (ii) there is (some sort of) resemblance between X and Y (formal criterion).

Aristotle, unlike Dretske and Millikan, would not be satisfied with a purely naturalist/physicalist explanation of representation. This, for Aristotle, could only be a part of the truth about the story of representation. One can think of the case of a ring leaving its sign on wax, and this is just a material process. Similarly, when air picks up a smell, this will be analyzed as a material process. In the case of sensation, Aristotle also talks about the material process that holds between a sense organ and sense-object. But, he is also aware that this would not be enough to understand the nature of sensation. After his discussion of the case of air receiving sensible form, he poses the following question: "What then is to smell, save to be 'affected somehow'? But to smell is to sense. Air, however, being so affected, becomes rapidly sensible".⁴¹⁶ If sensation were to be analyzed only from a material perspective, we could say nothing beyond the mere cause-effect duplication of the sensible form. The crucial idea here seems to be the following one: "a thing moved

⁴¹⁶ Aristotle, *De Anima*, 424b16-18. Here I use the translation of the text in St. Thomas Aquinas. *Commentary on Aristotle's De Anima*, translated by Kenelm Foster and Silvester Humphries (Notre Dame, Indiana: Dumb Ox Books, c1994), p. 171.

may move another” and “that which, being moved, moves another, must cause a motion similar to its own”.⁴¹⁷ So, if sensation were to be explained only as a physical movement of a sensible form from one object to another, then a sense organ, as being also a physical patient object, would need to cause a similar motion in such a way to become something sensible, like in the case of air, and this material process would go *ad infinitum*.⁴¹⁸ In the middle of this mechanic/material explanation, Aristotle would have said, we lose something. To smell, after all, is to sense. This cannot be captured only through a physical analysis. So, in addition to the naturalist/physicalistic criterion, Aristotle is also trying to give a formal condition/criterion in *De Anima* for sensation. The same must also be true of his account of representation, what is being represented in my mind at the personal level. This is very crucial, I believe, for understanding exactly what Aristotle tries to achieve in *De Anima*. The formal condition/criterion, which is the resemblance relation, is a very specific, sort of *sui generis* relation that exists only between mind and world.⁴¹⁹ That is to say, the resemblance relation between mind and world is not a physical/material process that can be found in nature, be it literal or non-literal resemblance. There is not any sort of resemblance relation in the world where the resembling object becomes one with the resembled. Aristotle, in some places in *De Anima*, says that when a subject, for instance, thinks of horse-ness, his mind becomes one with the

⁴¹⁷ St. Thomas Aquinas. *Commentary on Aristotle's De Anima*, translated by Kenelm Foster and Silvester Humphries (Notre Dame, Indiana: Dumb Ox Books, c1994), p. 202.

⁴¹⁸ This is my way of reading the whole discussion in *De Anima*, 424a17-424b20.

⁴¹⁹ I owe this point to Stephen Voss.

intellectual form of the object of his thought, horse-ness in the object.⁴²⁰

So, in a way, Aristotle actually tells us that the formal criterion for representation is to have the idea/form of the object in nature, this is what determines what is being represented in the mind. If you do not have the idea/form “horse-ness”, and instead if you somehow have the idea/form “mule-ness” and think of this as the form of horse, then you are actually misrepresenting a mule as a horse.⁴²¹ That is to say, if a subject does not understand the essence/form “horse-ness” and thinks that a mule is a horse, then his representation will be incorrect. The formal criterion here is actually a metaphysically given criterion. So, Forms/Ideas of objects are actually given in nature, but it is up to the individual human being to discover those forms in nature by some sort of appropriate investigation. If the investigation/understanding falls short of representing the essence/form of the object, then, Aristotle would say, the subject is in a case of misrepresenting the world/nature.⁴²²

Let’s summarize some of the crucial points in Aristotle’s account. The fact that mind is part of nature, for Aristotle, does not mean that representation consists in a physical/material process.⁴²³ Just giving a material criterion for representation, like Dretske and Millikan do, is just not enough for understanding the nature of mind and representation. We should also look for a formal criterion for what is being represented and what is being misrepresented in the personal level. This is very crucial.

⁴²⁰ For this point, see Chapter 2.

⁴²¹ I owe this example to Chryssi Sidiropoulou.

⁴²² I owe this point to Chryssi Sidiropoulou.

⁴²³ I owe this point to Chryssi Sidiropoulou.

Secondly, the formal criterion, albeit a very difficult criterion to grasp, in Aristotle's account actually gives us a conceptual condition for the cases of representation and misrepresentation. This is not the sort of explanation that tells us how, for instance, misrepresentation can occur, rather it tells us exactly the conceptual analysis of misrepresentation in a theory, whether the theory makes room for misrepresentation, something, for instance, that is not found in Dretske's story of misrepresentation.

I believe that the most accurate and fruitful account so far is Aristotle's account of mind. First of all, we have a functionalist perspective that divorces itself from any Platonic or Cartesian intuitions, and thereby it de-mystifies the nature of mind. Secondly, it tells us something very important about the direction of investigation concerning the nature of mind and representation. The philosophical investigation should not only limit itself to the physical / material explanation, it should go further and give a formal / conceptual analysis of what is being represented and misrepresented in mind. This is very demanding and it really pushes one to practice philosophy as excellently as possible. Following Aristotle's rigorous practice of philosophy, I think we should take the same attitude when we deal with the very nature of mind as a representing capacity. Representation is a very complex intellectual capacity, and we just cannot explain it through a materialist functionalist method, the method that is employed by Dretske and Millikan. If you take representation in isolation from other mental states and try to give a materialist / functionalist description of representation, you will definitely end up with a sort of difficulty in explaining the ridiculous cases such as

air representing wind, water representing the temperature, and you will not ever come close to understanding the very nature of mental representation. These sorts of causal theories always forget to mention the formal aspect of representation as in the case of a physicist who forgets to mention that anger is partly constituted by a desire to harm. Likewise, representing is partly constituted by, among other things, desiring to find a way back home, hoping to understand what other human beings represent, abstracting the object of thought from its particular features in a space-time continuum, associating different sorts of objects under a new class, writing a story about one's experience of the world. Representation, as Aristotle would say, is essentially / conceptually related to both perceiving and thinking. That is the nature of representation, which very much resists any sort of material reduction.

CHAPTER VIII

CONCLUSION

The purpose of this chapter is (i) to summarize some of the critical points of this thesis, and (ii) look for the implications of those critical points for further philosophical studies concerning the nature of mind as a representing capacity.

As the outcome of my research on the idea of mental representation, I have found that there exist, to my knowledge, two different approaches in the history of philosophy. On the one hand, people like Wittgenstein and Ryle accept the obvious fact that part of being a subject of mind is to be able to represent the world. This does not automatically force one to accept a more-loaded claim that there are mental representations in the mind. So, the second group of people, not satisfied only with the less-loaded claim, go further and claim that part of being a subject is to have mental representations in the mind, structures that establish one's sensory and cognitive relationship with the world outside. Leaving aside their different approaches to the idea of representation, they, at least, hold the same view, the view that can be traced back to Aristotle: Having a mind is, at bottom, a representational capacity.

If one holds the second approach to be true, namely, mental representations are those structures in the mind as the explanation for the representing mind, then he/she is faced with one of the most perplexing questions in the history of philosophy: what is the ground of mental representation? In virtue of what does something in the mind come to

represent something else in the world? In this context, the primary target of this thesis is to find an answer to this question by surveying some of the theories of mental representation in the history of philosophy.

In the second chapter, I presented and examined Aristotle's account of representation in *De Anima*. Aristotle's theory of mind is important for two reasons: (i) first, it is one of the earliest sources for the idea of mental representation, and it proves to be useful for understanding the underlying motives for Aristotle's commitment to the presence of mental representations in the mind; (ii) second, Aristotle's account of representation has not been examined exhaustively in the contemporary literature on mental representation. I believe that the literature missed something important here, because, as far as I understand, Aristotle provides a different answer for the question about the ground of mental representations, one that has not been even considered in the contemporary literature on mental representation. Aristotle, roughly, makes the following claim: for anything to represent something else, there must be a causation and (some kind of) resemblance between the representing vehicle and the represented. In the end of the chapter, I also talked about Aristotle's explanation for the occurrences of misrepresentation. I believe that one of the contributions of this thesis to the current literature is to remind people of Aristotle's account of representation.

Of course, so much has been changed in the twentieth century philosophy of mind since Aristotle. So, in order to understand the mainstream literature on mental representation, especially between 1980s and 1990s, one needs to be clear about the underlying metaphysical

framework at the background, which is, most often, a functionalist metaphysics of mind. However, it is hard to describe functionalism without making a comparison to previous popular metaphysics of mind such as behaviorism and type physicalism before the debut of functionalism. For this reason, in Chapter 3, I evaluate and critically examine three different theories of mind (behaviorism, type physicalism and functionalism) in the twentieth century philosophy of mind, each of which offers radically different accounts for the metaphysical questions like the following: “what is it for a subject to represent something else?” “In general, what is it to have a mind / mental states?” For behaviorism, mind is, in essence, a behavioral kind. Type physicalism, rejecting the acausal account of behaviorism, explains mind in terms of the causal power of neural states mediating between a set of stimuli and behavioral responses. Functionalism, on the other hand, rejects type-identifying mental states with neural states and offers a very liberal account of mind, according to which mind is identified as a functional kind (of appropriate complexity). In the last section of this chapter, I maintained the idea that contemporary functionalism, the one advertised by Putnam, actually rests on a very Platonic metaphysics, and I attempted to show how this Platonic metaphysics manifests itself through the most popular example of functionalist thought, that is, the software / hardware distinction. Then, I explained the reason why such a Platonic reading of functionalism can be very problematic, and, I suggested that the Aristotelian way of interpreting functionalism does seem to be a better philosophical venue for a common-sensical theory of mind. According to this interpretation, mind can be housed only within a limited range of physical / spiritual

structures, and the function and matter (physical/spiritual stuff) of mental states, contrary to the often-applied Platonic tendencies one may have, are very much connected. This seems to me a better way of holding functionalism. The whole discussion about the Platonic metaphysics behind contemporary functionalism, I believe, has some merit on its own. In general, people often think that contemporary functionalism has some connections with Aristotle's account of mind in *De Anima*. Contrary to the received view, I have found that contemporary functionalism actually rests on a very Platonic metaphysics, something that Aristotle wanted to abolish. I believe that this diagnosis is another contribution of this thesis to the literature.

As I said above, it is vitally important to have an idea about what functionalism is before examining theories of mental representation between the 1980s and 1990s, because most philosophers have become functionalists after Putnam's essay "Psychological Predicates". So, in order to understand, for instance, Dretske and Millikan's theories of mental representation in later chapters, one needs to know beforehand something about functionalism. But, introducing functionalism was not enough. Dretske and Millikan's theories of mental representation emerge as post-functionalist attempts to explain the individuation of content/mental representation. For this reason, in Chapter 4, I explained the transition from early functionalist programmes to post-functionalist programmes. In Chapter 4, I especially attempted to underline Putnam's change of mind in "The Meaning of 'Meaning'" and how the trouble with early versions of functionalism gives rise to post-functionalist programmes such as Dretske's indication theory and Millikan's

teleosemantic theory. The whole chapter is important to the extent that it prepares the groundwork for the later chapters.

In Chapter 5, I talked about Dretske's theory of representation in his pioneering work *Knowledge and the Flow of Information*, as one of those post-functionalist programmes. According to Dretske's view, mental representations are species of natural representations. So, understanding how natural representations represent will be the initial step towards understanding the nature of mental representations. That's the reason why Dretske begins with Grice's idea of "natural meaning", and develops this idea in order to offer a philosophical theory of information/representation, according to which information/representation is grounded in law-like correlations between the source and receiver. Dark clouds represent rain because of law-like correlations between dark clouds and rain. Smoke represents fire because of law-like correlations between smoke and fire. Likewise, representing vehicles in the mind represent something as F because of the law-like correlations between those representing vehicles and the objects with F. After introducing Dretske's view, in the end of chapter, I also dealt with the question whether Dretske's theory can solve the problem of misrepresentation.

In Chapter 6, I talked about the other post-functionalist attempt, Millikan's account in her famous article "Biosemantics", to explain the ground of a representation relationship between a representing vehicle in the mind and the represented object in the environment. First, I explained Millikan's criticism of Dretske's theory of representation. Then, I began introducing Millikan's way of explaining mental representation. The

most important concept for Millikan is “proper function”. A defective heart, for instance, may not perform its proper function, but we still say that pumping blood is its proper function. On the other hand, a molecular duplicate of me that emerges out of cosmic accident, like in the case of Davidson’s swampman, can have mechanisms that are identical with the mechanisms of a biological system, but, Millikan believes, the molecular duplicate would not have the right history, and therefore, its mechanisms (a heart, a kidney, an eye or a brain) would not have a proper function. If something does not have a proper function, then it is not a biological category. This is very crucial for Millikan, because she thinks that mentality, at bottom, is a biological phenomenon. In the end of this chapter, I again dealt with the question whether Millikan’s theory can explain the possibility of misrepresentation. It does seem that, once you agree with the basics of Millikan’s theory, misrepresentation is no longer a problem.

In Chapter 7, I attempted to wrap up the whole discussion in this thesis within the context of the problem of mental representation. Having provided enough material about the journey of the idea of mental representation in the history of philosophy, I wanted to discuss the virtues and vices of the theories of mental representation presented in the thesis. As far as I know, contrary to Fodor’s conviction, there have been three main options as an answer for the ground of mental representation, which are:

(R1) X represents Y iff there is (some sort of) resemblance between X and Y.

(R2) X represents Y iff there is a causal relationship between X and Y.

(R3) X represents Y iff (i) there is a causal relationship between X and Y, and (ii) there is (some sort of) resemblance between X and Y.

Millikan's teleosemantic theory of mental representation seems to be a better theory than Dretske's information theory of representation. However, this is not to say that it is a trouble-free theory of mind. In the second section of this chapter, I talk about the trouble with the contemporary literature on theories of mental representation in general and teleosemantics in particular. One of the working hypotheses behind the research strategies for probably most of the theories of mental representation in the late twentieth century was that any theory of mental representation must begin with the analysis of simple mental states/capacities and if the theory in question achieves a considerable success in explaining those simple ones, it will deal with complex mental states and capacities later.

As an offspring of this working hypothesis, most of the debates in this era revolve around the questions about fixing the content of simple representational states. As far as I can see, the literature got stuck with the first step (explaining simple mental states/capacities) and has not even managed to move to the next step (explaining complex mental states/capacities). A quick look at the debates about the frog's representation of a fly is only one instance indicating the trouble with the literature. I believe it is now just the right time to take the next task and start considering the questions like, for instance, whether there can be teleosemantic explanation for the complex representational states, for instance, "The Arab Spring has been a quite influential movement in the Middle East". It seems to me that cases of this sort are outside the

territory of teleosemantic explanation. So, while it may have been true of teleosemantic explanation that it can give an account of simple mental states, the same does not hold to be true for complex mental states. I don't see any way to develop teleosemantic theory to give an account of complex mental states, and therefore I abandon it.

In the last section of Chapter 7, I revisit Aristotle's account. Aristotle's account differs from the post-functionalist accounts defended by Dretske and Millikan, because, unlike Dretske and Millikan, it is not built on a Platonic metaphysics. Furthermore, there is a very crucial point in Aristotle, which, I believe, indicates exactly the missing perspective in the views of Dretske and Millikan. In this section, first I will summarize some of the main points in Aristotle's account of mind as a representing capacity, and then I will focus on his distinction between philosophical definition and physical/natural definition. I believe that understanding this difference helps us understand the good direction towards understanding the nature of mental representation, and show the source of the chronic problem within post-functionalist naturalist attempts to understand mind and representation. Following Aristotle, I believe that the fact that mind is part of nature does not mean that representation, as an essential aspect of mind, consists in a physical/material process. In understanding the nature of mind as a representing capacity, we should also look for the formal criterion for representation as Aristotle does in *De Anima*. The philosophical investigation should not only limit itself to the physical/material explanation, it should go further and give a formal/conceptual analysis of what is being represented and misrepresented in mind. This is very demanding and it really pushes one

to practice philosophy as excellently as possible. Following Aristotle's rigorous practice of philosophy, I think we should take the same attitude when we deal with the very nature of mind as a representing capacity. Representation is a very complex intellectual capacity, and we just cannot explain it through a materialist functionalist method, the method that is employed by Dretske and Millikan. If you take representation in isolation from other mental states and try to give a materialist/functionalist description of representation, you will definitely end up with a sort of difficulty in explaining the ridiculous cases such as air representing wind, water representing the temperature, and you will not ever come close to understanding the very nature of mental representation. These sorts of causal theories always forget to mention the formal aspect of representation as in the case of a physicist in *De Anima* who forgets to mention that anger is partly constituted by a desire to harm. Likewise, representing is partly constituted by, among other things, desiring to find a way back home, hoping to understand what other human beings represent, abstracting the object of thought from its particular features in a space-time continuum, associating different sorts of objects under a new class, writing a story about one's experience of the world. Representation, as Aristotle would say, is essentially / conceptually related to both perceiving and thinking. That is the nature of representation, which very much resists any sort of material reduction.

Now, I would like to talk about some of the implications that follow from my discussion about theories of mental representation in Chapter 7. So, let's start. First of all, it seems to me that at one point we really need to appeal to social, linguistic and cultural practices

surrounding mental beings. The very idea of naturalizing the mind, after this point in my research, sounds like a perfect dream that may never happen to become real. I have started to think that the idea of giving a naturalistic necessary and sufficient condition for a representation relationship between the mind and the world falls short of explaining and fully capturing the nature of mind as a representing capacity, because it ignores the crucial role social, linguistic and cultural interaction plays in shaping and determining one's representational power in conceiving the world and himself. I am sitting in a room now, writing these words, looking at various objects in my environment, and pose the following question: when I look at some objects in my room (e.g., a pencil), what is really being represented in my mind? As an answer to this question, I say "my mind represents this object as a pencil, not because there is some natural regularity between the object and mind, nor because there is some teleological/ evolutionary background behind my representation; but because I am part of this society and, in this society, a pencil is an important tool, and I just learnt how to use this tool, among other things, in order to express my ideas, feelings, etc. Actually, most of the things represented by my mind inherit their origin of representation from the rules and conventions of society." So, I say, instead of focusing only on the "natural" connection between the mind and world, we must also take social, linguistic and cultural exchange between the members of the society of minds into account. This point, to my knowledge, has not been sufficiently explored in the literature on mental representations. Giving attention to this aspect of the mind, that is, a functional structure constantly affecting other minds and being

affected by other minds in a social/ cultural setting, may open a completely new and promising research area. At one point, we can come to the point where non-reductive accounts of representation may, for instance, prove to be more adequate in dealing with complex capacities than reductive accounts of representation.

Focusing on the social, cultural practices may not only open a new research field, it may, on my view, give us a more accurate way to understand mind as a representing capacity. Understanding the nature of mind at its fullest would then certainly help us to resolve some of the debates, for instance, about artificial psychology, whether there can be systems that possess all the genuine and real properties of intelligence and cognition present in human beings and animals except the property of being “natural”. After understanding the mind as a socially encapsulated representation system, we can, for instance, test whether artificially intelligent systems can develop their mental skills in an appropriate society of minds.

WORKS CITED

- Adams, Fred and Aizawa, Ken. Spring 2010. Causal Theories of Mental Content. Available [online]:
"http://plato.stanford.edu/archives/spr2010/entries/content-causal/" [Spring 2010].
- Andersen, Lise Marie. *Functionalism and Embodied, Embedded Mind – The Extended Story*. M.A. Thesis, Edinburg Research Archive, 2007.
- Anscombe, Gertrude Elizabeth Margaret. *Intention*. Cambridge, Mass.: Harvard University Press, 2000.
- Aristotle. *On The Soul; Parva Naturalia; On Breath*, translated by W. S. Hett. Cambridge, Mass.: Harvard University Press, 2000, c1957.
- Armstrong, David. "The Nature of Mind." In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block. Cambridge, Mass.: Harvard University Press, 1980a.
- Armstrong, David. *A Materialistic Theory of the Mind*. London: RKP, 1968.
- Armstrong, David. "The Causal Theory of Mind." In *The Nature of Mind*, edited by David M. Rosenthal. New York: Oxford University Press, c1991.
- Aquinas, St. Thomas. *Commentary on Aristotle's De Anima*, translated by Kenelm Foster and Silvester Humphries. Notre Dame, Indiana: Dumb Ox Books, c1994.
- Ben-Menahem, Yemima (Ed). *Hilary Putnam (Contemporary Philosophy in Focus)*. Cambridge, New York: Cambridge University Press, 2005.
- Bird, Alexander. *Nature's Metaphysics: Laws and Properties*. Oxford: Clarendon Press; New York: Oxford University Press, 2007.
- Block, Ned (Ed.). *Readings in Philosophy of Psychology*. Vol. 1. Cambridge, Mass.: Harvard University Press, 1980.
- Block, Ned. "Introduction: What is Functionalism?" In *Readings in Philosophy of Psychology*, edited by Ned Block. Cambridge, Mass.: Harvard University Press, 1980a.
- Block, Ned. "The Computer Model of the Mind." In *Readings in Philosophy and Cognitive Science*, edited by Alvin I. Goldman. Cambridge, Mass.: Harvard University Press, 1993.

- Block, Ned. "On a Confusion about the Function of Consciousness." *Behavioral and Brain Sciences* 18, 2 (1995), pp. 227-247.
- Brown, Deborah. "A Furry Tile About Mental Representation." *The Philosophical Quarterly* (October 1996), pp. 448-466.
- Burnyeat, M. F. "Is an Aristotelian Philosophy of Mind Still Credible? (A Draft)." In *Essays on Aristotle's De Anima*, edited by Martha C. Nussbaum and Amelie Oksenberg Rorty. Oxford [England]: Clarendon Press, 1997.
- Campbell, Keith. *Body and Mind*. Notre Dame, Ind.: University of Notre Dame Press, 1984.
- Capitan, W. H. and Merrill, D. D. (Eds.). *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press, 1967.
- Clapin, Hugh (Ed.). *Philosophy of Mental Representation*. New York, Oxford: Oxford University Press, 2002.
- Clapin, Hugh. "Introduction." In *Philosophy of Mental Representation*, edited by Hugh Clapin. New York, Oxford: Oxford University Press, 2002.
- Clapin, Hugh; Staines, Phillip and Slezak, Peter (Eds.). *Representation in Mind: New Approaches to Mental Representation*. Amsterdam; Boston: Elsevier, 2004.
- Copeland, B. Jack. *Artificial Intelligence: A Philosophical Introduction*. Cambridge: Blackwell, 1995.
- Craig, Edward (Ed.). *Routledge Encyclopedia of Philosophy*. London, New York: Routledge, 1996, vol. 4.
- Crane, Tim. "Intentionality." In *Routledge Encyclopedia of Philosophy*, edited by Edward Craig. London, New York: Routledge, 1996, vol. 4.
- Crane, Tim. *The Mechanical Mind*. London, New York: Routledge, 2003.
- Cummins, Robert. *Meaning and Mental Representation*. Cambridge, London: MIT Press, 1989.
- Cummins, Robert and Poirier, Pierre. "Representation and Indication." In *Representation in Mind: New Approaches to Mental Representation*, edited by Hugh Clapin, Phillip Staines, Peter Slezak. Amsterdam; Boston: Elsevier, 2004.

- Davidson, Donald. *Essays on Actions and Events*. Oxford: Clarendon Press; New York: Oxford University Press, 1980.
- Davidson, Donald. "Actions, Reasons and Causes." In *Essays on Actions and Events*. Oxford: Clarendon Press; New York: Oxford University Press, 1980.
- Dawkins, Richard. *The Blind Watchmaker*. New York: Norton, 1986.
- Dennett, Daniel C. *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, Mass.: MIT Press, 1978.
- Dennett, Daniel C. "A Cure for the Common Code?" In *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, Mass.: MIT Press, 1978.
- Dennett, Daniel C. "Artificial Intelligence as Philosophy and as Psychology." In *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, Mass.: MIT Press, 1978.
- Dennett, Daniel C. *The Intentional Stance*. Cambridge, London: MIT Press, 1987.
- Dennett, Daniel C. "Styles of Mental Representation." In *The Intentional Stance*. Cambridge, London: MIT Press, 1987.
- Dennett, Daniel C. "Intentional Systems in Cognitive Ethology: The 'Panglossian Paradigm' Defended." In *The Intentional Stance*. Cambridge, London: MIT Press, 1987.
- Dretske, Fred I. *Seeing and Knowing*. London: Routledge & K. Paul, 1969.
- Dretske, Fred I. *Knowledge and the Flow of Information*. Stanford, CA: CSLI Publications, 1999; originally published: Cambridge, Mass.: MIT Press, c1981.
- Dretske, Fred I. "Précis of Knowledge and the Flow of Information." *Behavioral and Brain Sciences* 6, no. 1 (1983), pp. 55-63.
- Dretske, Fred I. *Explaining Behavior: Reasons in a world of causes*. Cambridge, Mass.: MIT Press, c1988.
- Dretske, Fred I. "Misrepresentation." In *Mental Representation: A Reader*, edited by Stephen P. Stich and Ted Warfield. Oxford, UK; Cambridge, USA: Blackwell, 1994.
- Dretske, Fred I. *Naturalizing the Mind*. Cambridge, Mass.: MIT Press, c1995.

- Dretske, Fred I. *Perception, Knowledge and Belief*. Cambridge, U.K.; New York: Cambridge University Press, 2000.
- Eliasmith, Chris. "The Myth of the Turing Machine: The Failings of Functionalism and Related Theses." *Journal of Experimental and Theoretical Artificial Intelligence* 14, no. 1 (February 2002), pp. 1-8.
- Eliasmith, Chris, "An Introduction to Contemporary Theories of Content" [Unpublished]. Available [online]:
http://www.arts.uwaterloo.ca/~celiasmi/courses/old_courses/Fall01/Phil600Repn/papers/eliasmith.2001.pdf
- Fodor, Jerry A. "The Mind-Body Problem." *Scientific American* 244 (1981), pp. 114-125.
- Fodor, Jerry A. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press, 1987.
- Fodor, Jerry A. *A Theory of Content and Other Essays*. Cambridge, Mass.: MIT Press, 1992.
- Fodor, Jerry A. "Semantics, Wisconsin Style." In *A Theory of Content and Other Essays*. Cambridge, Mass.: MIT Press, c1992.
- Fodor, Jerry A. "Meaning and the World Order." In *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press, 1987.
- Fodor, Jerry A. "Deconstructing Dennett's Darwin." *Mind and Language* 11, no. 3 (1996), pp. 246- 262.
- Fodor, Jerry A. *The Elm and the Expert: Mentalese and its Semantics*. Cambridge, Mass.: MIT Press, 1995.
- Frege, Gottlob. *Aritmetiğin Temelleri*, translated by Bülent Gözkan. Istanbul: Yapı Kredi Yayınları, 2008.
- Gallistel, C. R. *The Organisation of Learning*. Cambridge, Mass.: MIT Press, 1960.
- Godfrey-Smith, Peter. *Complexity and the Function of Mind in Nature*. Cambridge; New York: Cambridge University Press, c1998.
- Goldman, Alvin I. (Ed.). *Readings in Philosophy and Cognitive Science*. Cambridge, Mass.: Harvard University Press, 1993.
- Goodman, Nelson. *Languages of Art: An Approach to a Theory of Symbols*. London: Oxford University Press, 1969.

- Gözkan, Bülent. "Çevirenin Sunuşu." In *Gottlob Frege, Aritmetiğin Temelleri*, translated by Bülent Gözkan, 2008.
- Green, Christopher D. "The Thoroughly Modern Aristotle: Was Aristotle really a functionalist?" *History of Psychology* 1, no. 8 (February 1998), pp. 8-20.
- Grice, Paul. "Meaning." *Philosophical Review* 66, no. 3 (Jul., 1957), pp. 377-388.
- Griffiths, Paul E. "Functional Analysis and Proper Functions." *British Journal for the Philosophy of Science* 44, no. 3 (1993), pp. 409-422.
- Griffiths, Paul E. & Goode, R. "The Misuse of Sober's Selection for/Selection of Distinction." *Biology and Philosophy* 10, no. 1(1995), pp. 99-107.
- Gunderson, K (Ed.). *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science*, Vol. 7. Minneapolis: University of Minnesota Press, 1975.
- Guttenplan, Samuel (Ed.). *A Companion to the Philosophy of Mind*. Cambridge: Blackwell, 1994.
- Heil, John. *Philosophy of Mind: A Contemporary Introduction*. London, New York: Routledge, 1998.
- Heil, John. "Multiple Realizability." *American Philosophical Quarterly* 36, no. 3 (July 1999), pp. 189-208.
- Hook, Sidney (Ed.). *Dimensions of Mind*. New York: New York University Press, 1960.
- Irmak, Nurmay (forthcoming), "Software is an Abstract Artifact." *Grazer Philosophische Studien*.
- Jackson, Frank. "Representation and Experience." In *Representation in Mind: New Approaches to Mental Representation*, edited by Hugh Clapin, Phillip Staines, Peter Slezak, 2004.
- Jaworski, William. 24 February 2011. *Mind and Multiple Realizability*. Available [online]: "<http://www.iep.utm.edu/mult-rea/>" [24 February 2011].
- Johnson, David Martel and Erneling, Christina E. (Ed.). *The Future of the Cognitive Revolution*. New York, Oxford: Oxford University Press, 1997.

- Kahn, Charles H. "Aristotle on Thinking." In *Essays on Aristotle's De Anima*, edited by Martha C. Nussbaum and Amelie Oksenberg Rorty. Oxford [England]: Clarendon Press, 1997.
- Kant, Immanuel. *Kant: Philosophical Correspondence, 1759-99*, edited and translated by Arnulf Zweig. Chicago: Chicago University Press, 1967.
- Kazantzakis, Nikos. *Zorba the Greek*, translated by Carl Wildman. London: Faber and Faber, 1961.
- Kim, Jaegwon. *Philosophy of Mind*. Boulder, Col.: Westview Press, 1996.
- Kingsbury, Justine. "A Proper Understanding of Millikan." *Acta Analytica* 21, no. 3 (2006), pp. 23-40.
- Kosslyn, Stephen M. and Hatfield, Gary. "Representation Without Symbol Systems." *Social Research* 51, no. 4 (Winter 1984), pp. 1019-1045.
- Lagerlund, Henrik. Fall 2008. *Mental Representation in Medieval Philosophy*. Available [online]=
["http://plato.stanford.edu/archives/fall2008/entries/representation-medieval/"](http://plato.stanford.edu/archives/fall2008/entries/representation-medieval/) [Fall 2008].
- Le Guin, Ursula K. *The Left Hand of Darkness*. New York: Ace Books, 2000, c1969.
- Le Guin, Ursula K. "Is Gender Necessary? Redux." In *Dancing at the Edge of the World: Thoughts on Words, Women, Places*. New York: Groove Press, 1989.
- Lewis, David. "Review of Putnam." In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block. Cambridge, Mass.: Harvard University Press, 1980a.
- Lewis, David. "Psychophysical and Theoretical Identifications." In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block. Cambridge, Mass.: Harvard University Press, 1980a.
- Lewis, David. "An Argument for the Identity Theory." *Journal of Philosophy* 63, no. 1 (January, 1966), pp. 17-25.
- Loewer, Brian & Rey, George (Eds.). *Meaning in Mind: Fodor and His Critics*. Cambridge, MA: Blackwell, 1991.
- Lycan, William G. (Ed.). *Mind and Cognition: A Reader*. Cambridge, Mass., USA: Basil Blackwell, 1990.

- Lycan, William G. "The Continuity of Levels of Nature." In *Mind and Cognition: A Reader*, edited by William G. Lycan. Cambridge, Mass., USA: Basil Blackwell, 1990.
- Macdonald, Graham and Papineau, David (Eds.). *Telesemantics*. Oxford, UK: Clarendon Press, 2006.
- Mann, Thomas. *Transposed Heads: A legend of India*. Translated by H. T. Lowe-Porter. New York: A. A. Knopf, 1941.
- McLaughlin, Brian; Beckermann, Ansgar and Walter, Sven (Eds.). *The Oxford Handbook of Philosophy of Mind*. Oxford: Clarendon Press; New York: Oxford University Press, 2009.
- Millikan, Ruth. "Biosemantics." *Journal of Philosophy* 86, no. 6 (June 1989), pp. 281-297.
- Millikan, Ruth. *Language, Thought and Other Biological Categories: New Foundations for Realism*. Cambridge, Mass.; London, England: MIT Press, 1984.
- Millikan, Ruth. "In Defense of Proper Functions." *Philosophy of Science* 52, no. 2 (June 1989), pp. 288-302.
- Millikan, Ruth. "Truth Rules, Hoverflies and the Kripke-Wittgenstein Paradox." *Philosophical Review* 99 (1990), pp. 232-53.
- Millikan, Ruth. "Truth Rules, Hoverflies and the Kripke-Wittgenstein Paradox." In *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press, 1993.
- Millikan, Ruth. *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press, 1993.
- Millikan, Ruth. "Speaking Up for Darwin." In *Meaning in Mind: Fodor and His Critics*, edited by Loewer, B. & Rey, G. Cambridge, MA: Blackwell, 1991.
- Millikan, Ruth. "Biosemantics." In *The Oxford Handbook of Philosophy of Mind*, edited by Brian McLaughlin; Ansgar Beckermann and Sven Walter. Oxford: Clarendon Press; New York: Oxford University Press, 2009.
- Modrak, Deborah K. W. "The Nous-Body Problem in Aristotle." *Review of Metaphysics* 44, no. 4 (June 1991), pp. 755-774.
- Moor, James H. "Three Myths of Computer Science." *The British Journal for the Philosophy of Science* 29, no. 3 (September, 1978), pp. 213-222.

- Moore, Robert. "A Cognitivist Reply to Behaviorism." *The Behavioral and Brain Sciences* 7, no. 4 (December 1984), pp. 637-639.
- Neander, Karen. "Malfunctioning and Misrepresenting." *Philosophical Studies* 79, no. 2 (1995), pp. 109-141.
- Neander, Karen. Spring 2012. Teleological Theories of Mental Content. Available [online]:
 "http://plato.stanford.edu/archives/spr2012/entries/content-teleological/[Spring 2012]".
- Nussbaum, Martha C. and Rorty, Amelie Oksenberg (Eds.). *Essays on Aristotle's De Anima*. Oxford [England]: Clarendon Press, 1997.
- Nussbaum, Martha C. and Putnam, Hilary. "Changing Aristotle's Mind." In *Essays on Aristotle's De Anima*, edited by Martha C. Nussbaum and Amelie Oksenberg Rorty. Oxford [England]: Clarendon Press, 1997.
- O'Brien, Gerard and Opie, Jon. "Towards a Structural Theory of Mental Representation." In *Representation in Mind: New Approaches to Mental Representation*, edited by Hugh Clapin, Phillip Staines, Peter Slezak. Amsterdam; Boston: Elsevier, 2004.
- Papineau, David. "Representation and Explanation." *Philosophy of Science* 51, 4 (December, 1984), pp. 550-572.
- Papineau, David. *Reality and Representation*. Oxford UK, Cambridge USA: Blackwell, 1987.
- Papineau, David. *Philosophical Naturalism*. Oxford UK, Cambridge USA: Blackwell, 1993.
- Perner, Josef. *Understanding the Representational Mind*. Cambridge, Mass.; London, England: MIT Press, 1991.
- Pietroski, Paul M. "Intentional and Teleological Error." *Pacific Philosophical Quarterly* 73, no. 3 (1992), pp. 267-81.
- Puccetti, Roland. "The Great C-Fiber Myth: A Critical Note." *Philosophy of Science* 44, no. 2 (June 1977), pp. 303-305.
- Putnam, Hilary. "Minds and Machines." In *Dimensions of Mind*, edited by Sidney Hook. New York: New York University Press, 1960.
- Putnam, Hilary. "Psychological Predicates." In *Art, Mind, and Religion*, edited by W. H. Capitan & D. D. Merrill. Pittsburgh: University of Pittsburgh Press, 1967.

- Putnam, Hilary. "The Meaning of 'Meaning.'" In *Language, Mind and Knowledge*, edited by K. Gunderson. *Minnesota Studies in the Philosophy of Science*, Vol. 7. Minneapolis: University of Minnesota Press, 1975.
- Putnam, Hilary. "The Meaning of 'Meaning.'" In *Mind, Language and Reality*. Cambridge: Cambridge University Press, 1975.
- Putnam, Hilary. *Mind, Language and Reality*. Cambridge, New York: Cambridge University Press, 1975.
- Putnam, Hilary. "Brains and Behavior." In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block. Cambridge, Mass.: Harvard University Press, 1980a.
- Putnam, Hilary. "The Nature of Mental States." In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block. Cambridge, Mass.: Harvard University Press, 1980a.
- Putnam, Hilary. "Philosophy and Our Mental Life." In *Readings in Philosophy of Psychology*, Vol. 1, edited by Ned Block. Cambridge, Mass.: Harvard University Press, 1980a.
- Putnam, Hilary. *Reason, Truth and History*. Cambridge, New York: Cambridge University Press, 1981.
- Putnam, Hilary. *Representation and Reality*. Cambridge, Mass.: MIT Press, 1991, c1988.
- Putnam, Hilary. *Renewing Philosophy*. Cambridge, London: Harvard University Press, 1992.
- Putnam, Hilary. "Does Evolution Explain Representation?" In *Renewing Philosophy*. Cambridge, London: Harvard University Press, 1992.
- Putnam, Hilary. "Putnam, Hilary." In *A Companion to the Philosophy of Mind*, edited by Samuel Guttenplan. Cambridge: Blackwell, 1994.
- Putnam, Hilary. "Functionalism: Cognitive Science or Science Fiction?" In *The Future of the Cognitive Revolution*, edited by David Martel Johnson and Christina E. Erneling. New York, Oxford: Oxford University Press, 1997.
- Ramachandran, V. S. & Blakeslee, Sandra. *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. New York: William Morrow and Company, Inc., 1999.
- Robinson, Daniel N. *Aristotle's Psychology*. S. I: Joe Christensen Inc.: 1999.

- Rosenthal, David M. (Ed.). *The Nature of Mind*. New York: Oxford University Press, c1991.
- Rowlands, M. (1999). Teleosemantics. *Online A Field Guide to Philosophy of Mind*. Available [online]:
["http://host.uniroma3.it/progetti/kant/field/teleo.html"](http://host.uniroma3.it/progetti/kant/field/teleo.html).
- Ryle, Gilbert. *The Concept of Mind*. London: Penguin, 2000.
- Simon, H.A., and A. Newell. "Heuristic Problem Solving: The Next Advance in Operations Research." *Operations research* 6, no. 1 (1958), pp. 1-10.
- Shagrir, Oron. "The Rise and Fall of Computational Functionalism." In *Hilary Putnam (Contemporary Philosophy in Focus)*, edited by Yemima Ben-Menahem. Cambridge, New York: Cambridge University Press, 2005.
- Shields, Christopher. "The First Functionalist." In *Historical Foundations of Cognitive Science*, edited by J. C. Smith. Dordrecht: Kluwer Academic Publishers, 1991.
- Shields, Christopher. Winter 2008. *Aristotle's Psychology*. Available [online]=
["http://plato.stanford.edu/archives/win2008/entries/aristotle-psychology/"](http://plato.stanford.edu/archives/win2008/entries/aristotle-psychology/) [Winter 2008].
- Shields, Christopher. "Controversies surrounding Aristotle's Theory of Perception." In *Shields, Aristotle's Psychology*.
- Sidiropoulou, Chryssi (forthcoming). "A New Understanding of Aristotle's 'Dualism' in the De Anima." *Philosophical Inquiry*.
- Skinner, Burrhus Frederic. *Walden Two*: Hackett Publishing Company Incorporated, 2005.
- Skinner, Burrhus Frederic. *Science and Human Behavior*. New York: Free Press, 1965.
- Smart, J. J. C. "Sensations and Brain Processes." *Philosophical Review* 68 (1959), pp. 141-156; reprinted in *The Nature of Mind*, ed. David M. Rosenthal. New York: Oxford University Press, 1991.
- Smith, J. C (Ed.). *Historical Foundations of Cognitive Science*. Dordrecht: Kluwer Academic Publishers, 1991.
- Sorabji, Richard. "Body and Soul in Aristotle." *Philosophy* 49, no. 187 (January, 1974), pp. 63–89.

- Sperber, Dan. "Metarepresentation." In *The MIT Encyclopedia of Cognitive Sciences*, edited by Robert A. Wilson and Frank C. Keil. Cambridge Mass.; London, England: The MIT Press, 1999.
- Sperber, Dan (Ed.). *Metarepresentations: A Multidisciplinary Perspective*. Oxford, New York: Oxford University Press, 2000.
- Stampe, Dennis W. "Towards a Causal Theory of Linguistic Representation." *Midwest Studies in Philosophy* 2, no. 1 (September 1977), pp. 42-63.
- Sterelny, Kim. *The Representational Theory of Mind: An Introduction*. Oxford, Mass.: Blackwell, 1990.
- Sterelny, Kim. "Basic Minds," *Philosophical Perspectives* 9, AI, Connectionism and Philosophical Psychology (1995), pp. 251-270.
- Stich, Stephen P. and Warfield, Ted (Eds.). *Mental Representation: A Reader*. Oxford, UK; Cambridge, USA: Blackwell, 1994.
- Stevens, Michael. *The Explanatory Role of the Notion of Representation* [Unpublished]. Available online: "<http://www.stevens.org/research/expln/mimeoexpln.pdf>".
- Suber, Peter. "What is Software?" *The Journal of Speculative Philosophy* 2, No. 2 (1988), pp. 89-119.
- Von Eckardt, Barbara. *What is Cognitive Science?* Cambridge, Mass.: London, England: MIT Press, 1996.
- Watson, Richard A. *Representational ideas: From Plato to Patricia Churchland*. Dordrecht, Boston: Kluwer Academic Publishers, 1995.
- Watson, John B. "Psychology as the Behaviorist Views It." *Psychological Review* 20, no. 2 (1913), pp. 158-77.
- Whitehead, Alfred North and Russell, Bertrand. *Principia Mathematica*. Cambridge [Eng.]: University Press, 1960.
- Wilson, Robert A. and Keil, Frank C. (Eds.). *The MIT Encyclopedia of Cognitive Sciences*. Cambridge, Mass.; London, England: The MIT Press, 1999.
- Wilson, Robert A. "Philosophy." In *The MIT Encyclopedia of Cognitive Sciences*, edited by Robert A. Wilson and Frank C. Keil. Cambridge, Mass.; London, England: The MIT Press, 1999.

Wittgenstein, Ludwig. *Philosophical Investigations*. Oxford: Blackwell, 1991, c1953.