GROUNDING MENTAL CONTENT

by

Christine June Young

B.A. Cornell University, 2000

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

THE KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Christine June Young

It was defended on

November 28, 2011

and approved by

John McDowell, Distinguished University Professor of Philosophy,

Department of Philosophy

Michael Thompson, Professor, Department of Philosophy

Peter Machamer, Professor, Department of History and Philosophy of Science

Dissertation Advisor: Anil Gupta, Distinguished Professor, Department of Philosophy

GROUNDING MENTAL CONTENT

Christine J. Young, PhD

University of Pittsburgh, 2012

Beliefs and desires' functions as bases for action are essential for their contents. A belief can only be a basis for a particular action if it is about the object targeted by that action. Whether a belief can legitimately play that role for an action depends on its intentional relation with the world. However, one can have beliefs about objects which do not exist. The intentionality of empty beliefs cannot be construed as a relation between these beliefs and external objects. Yet without this belief-object connection, they can still function legitimately as bases for action. How is this possible?

What we need is an adequate account of how a belief's intentionality is grounded. But perhaps grounding does not have to take the form of an intentional relation between a token belief and its referent. It can be more indirect and complex, so that beliefs' functions as bases for action are not jeopardized by non-existence of their referents. One thing seems undeniable. One can only have beliefs about particular objects, if one interacts with the world. The question is: how are beliefs grounded in such interactions? This is the central question for my dissertation.

This essay has four chapters. In the first chapter, I consider the idea that all empty thoughts can be grounded in object-dependent thoughts in subject-world interactions. The arguments I will consider are Gareth Evans's in *Varieties of Reference*. I will argue that these arguments do not succeed. In the second section, I will consider a materialist proposal: thoughts are grounded in virtue of causal properties of subject-world interactions. For this section, I will

examine Dretske's functionalism in *Explaining Behavior*. I will argue that intentionality cannot arise from causal structures. In the third chapter, I will present my own positive view. The main ideas are: 1) intentionality of *token* beliefs or desires is grounded in the subject's representational *capacities*, 2) these capacities are grounded in how their development actualize the subject's intentional potential. The final chapter presents application of my positive views to contemporary debates.

TABLE OF CONTENTS

1.0 INTRODUCTION 1

1.1 MENTAL VIABILITY 2

1.2 REFORMULATING THE ORIGINAL PROBLEMS 3

2.0 OBJECT-DEPENDENT THOUGHTS 6

2.1 FREGE-RUSSELL BACKGROUND 7

2.1.1 Semantic value 7

2.1.2 Russell 10

2.1.3 Philosophy of Mind aspect of Russellian terms 10

2.1.4 Conclusion 11

2.2 OBJECT-DEPENDENCY 12

2.2.1 Information-based thought 12

2.2.2 Information 13

2.2.3 Thoughts and the Generality Constraint 14

2.2.4 Idea of an object 16

2.2.5 Interpretation and purpose 18

2.2.6 Object-dependent Idea-of-an-object 19

2.3 DEMONSTRATIVE THOUGHTS 20

2.3.1 Demonstrative Idea and demonstrate identification 21

2.3.2 Place identification 22
2.3.3 Demonstrative Ideas of objects 28
2.4 EVANS'S FRAMEWORK OF SUBJECT-WORLD INTERACTIONS 29
2.4.1 Methodology 30
2.4.2 Hallucination as thought. 31
2.4.3 Thought attribution and purpose 33
2.4.4 Coherence of thought content 37
2.4.5 Intelligibility of hallucinatory thoughts 38
3.0 DRETSKE'S MATERIALISM 45
3.1 THE NATURE OF BEHAVIOR 46
3.1.1 Movement vs. behavior. 46
3.1.2 Behavior as process. 48
3.1.3 Two kinds of causes. 50
3.2 REPRESENTATION AND FUNCTION 51
3.2.1 Indication 51
3.2.2 Function to indicate 54
3.2.3 Non-derived representational systems 56
3.2.4 Development 57
3.3 MEANING 59
3.3.1 Purposeful behavior 60
3.3.2 Learning as recruitment 63
3.3.3 Conclusion 64
3.4 MATERIALISM 66

3.4.1 The soprano effect 66
3.4.2 Modifiability and reflexivity 68
3.4.3 The materialist's challenge 69
3.5 INTENTIONAL EXPLANATIONS 71
3.5.1 Function, causation, and normativity 71
3.5.2 Indication and causation 72
3.5.3 Structural explanations vs. triggering explanations 73
3.5.4 Causation and explanation 73
3.6 INTENTIONAL EXPLANATIONS AND REPRESENTATIONAL CAPACITIES 74
3.6.1 Function and representational capacities 77
3.6.2 Accumulation of representational capacities 78
3.6.3 Subject-world interactions and normativity 79
3.7 NON-DERIVED NORMATIVITY 80
3.7.1 Learning and purpose 80
3.7.2 Subject's purpose 81
3.7.3 Two kinds of holism 85
3.7.4 Conclusion 87
4.0 A PROPOSAL 89
4.1 THE PROBLEM 89
4.1.1 Object of answerability 90
4.1.2 Capacity for thoughts 91
4.1.3 Acquisition of object-specific capacities 92

4.1.4 Conceptual holism 93
4.2 GOUNDING 96
4.2.1 Grounding vs. reduction 96
4.2.2 Grounding answerability 100
4.3 ANSWERABILITY 102
4.3.1 Passive answerability 102
4.3.2 The passive FP answerability problem 105
4.3.3 Active answerability 106
4.3.4 Active answerability from commitment 107
4.3.5 The structure of commitment 108
4.3.6 Dretske's structure 111
4.3.7 Summary 113
4.4 EXTERNAL WORLD AND EXTERNAL OBJECTS 115
4.4.1 The necessity for grounding the sense of the external 116
4.4.2 Distinction of self vs. the other 118
4.4.3 Movements 121
4.5 CONCLUSION 122
4.5.1 Summary 122
4.5.2 Object-independence 124
4.5.3 Reference and sense 126
4.5.4 Narrow vs. wide content 128
4.5.5 Direct and indirect representations 131
4.5.6 Concepts in general 134

5.0 CONCLUSION 136

5.1 PURPOSE 136

5.1.1 Purpose and the subject's perspective 137

5.1.2 Non-intentional phenomenal experiences. 138

5.1.3 Intentional explanations 139

5.1.4 Realism 140

5.1.5 Mind and world. 141

5.1.6 Structure of content 142

5.2 RATIONALITY 143

5.2.1 Concepts: representations invoking rationality 143

5.2.2 Rational and intentional frameworks. 146

5.2.3 Grounding rational intentionality 147

5.3 CONCLUSION 148

BIBLIOGRAPHY 150

LIST OF TABLES

Table 1. Aristotle's Potentiality and Actuality 100

LIST OF DIAGRAMS

Diagram 1. Dretske's Structure of Learning1 62

Diagram 2. Dretske's Structure for Tropistic Behavior 62

Diagram 3. Dretske's Structure of Learning2 112

1.0 INTRODUCTION

The intentionality relationship between thoughts and objects can seem elusive. It is not concrete the way causal relationships are. One can think about objects one is not in causal contact with, e.g. "My mother should be in Switzerland by now", "The Mayans had a sophisticated culture", "The Big Bang Theory is becoming popular." In fact, since at any one time, we are in causal contact only with a small number of things, most of our thoughts are not in causal contact with their referents. How are intentional relationships between thoughts and objects in the world (or states of affairs) established?

Matters become worse when we consider the possibility of empty thoughts. We can think about objects which do not exist. A child can wonder what Santa Claus is going to give her. In antiquity, people prayed to gods who are imaginary. These thoughts are flawed because their referents do not exist, not because they are not intentional. On the contrary, precisely because they are intentional, that is, they purport to be about something; not having a referent is a type of failure. (By contrast a pain state's not having a referent is not a failure of any sort.) In such cases, there is intentionality, but we cannot see intentionality as a relationship between a thought and a referent, since there is no referent. So here are some perplexing features of intentionality:

1) Intentionality seems to be a relationship between thought and its referent when the referent exists. But it is not impaired when the referent does not exist.

2) Even when the referent does exist, there is no concrete relationship which underpins this abstract relationship.

1.1 MENTAL VIABILITY

The problems of intentionality take on a different light when we turn to another aspect of thoughts: their connections with perceptions, desires, and action. These mental states[1] are connected by their referents. In acting on an object, the subject perceives, thinks about, desires, and develops intentions towards the same object. For a thought to have a viable role in the subject's mental world, its intentional relations must be connected with those of other mental states. If these mental states' roles in the subject's mental life are essential to their nature, then their intentionality cannot be fully in view when isolated from the rest of the network. The smallest unit of intentionality is not that of a thought, but that of a mind. Furthermore, since past and present thoughts are also connected, the fundamental intentionality is not just that of a mind at one point in time, but that of a mind with past, present, and future.

The holistic nature of intentionality reshapes the above problems. If we are looking for concrete connections underpinning intentionality, they would not be isolated causal connections between a thought and the object of the thought, but a set of causal connections between the subject's mind and the world the subject has been in contact with. It is the subject-world interactions over time which underpins the fundamental intentionality relationship.

[1] Unless otherwise qualified, I will use the phrase 'mental states' for 'intentional mental states.'

If a thought is about nothing, then the desire and the action which are based on it are also about nothing. Empty reference of one thought is not just a problem of that mental state alone, but spreads across a set of interconnected mental states. That it can spread in this way shows that these interconnections are immune to referential failure to a significant degree. So an empty thought can still generate other thoughts, beliefs, and desires, so forth. But they would all be contaminated by its referential failure.

If there are no thoughts which are immune to referential failure, and no check on the extension of such failure, then it seems possible for a mind to consist entirely of empty mental states. Given the historical nature of mind, the subject's entire mental history would be one massive hallucination, as though Descartes' evil genius had taken hold of it. The subject's mind would be completely cut off from the world. The intentionality of such an experience if possible would resemble a dream. In dreams one does not think or act on the external world. Even when one is sleep walking, one is not thinking about the actual destination, but the one in one's mind. These intentional states are not part of one's life in the real world. Though mental viability of thoughts (with the external world) is immune to localized hallucination, it requires that the fundamental intentional relation between mind and world be intact.

1.2 REFORMULATING THE ORIGINAL PROBLEMS

The holistic nature of intentionality helps us resolve the first initial problem that intentionality seems to be a relationship when the thought's referent exists, but one which is not impaired when the thought is empty. This seeming contradiction does not apply to the mind's intentional relation to the world. The possibility of holistic, massive hallucination is not a real threat.

Though localized hallucinations are very real, there is no intuition to support the possibility of such mass hallucination. So we do not have a pre-theoretical threat which our theory must deal with. On the contrary any claim about the possibility of mass hallucination would come from a philosophical theory. For our pre-theoretical stance, we can say that intentionality is a relationship between mind and world without worrying about the absence of one relatum.

The other problem remains: how can intentionality be underpinned by some concrete relationship? At the level of thoughts, these relationships are supposed to determine what the subject is thinking about and how she is thinking about it. Now in addition, these concrete relationships are needed to ensure that the intentional relation between mind and world remains intact despite occasional empty thoughts.

Minimally these concrete relationships will constitute causal interactions between the subject and the world. It is clear that these interactions do not ground thoughts one by one. How do they ensure that the intentional relationship between mind and world is intact, and how do they determine how the subject thinks about particular aspects of the world? This is the question for this essay.

To address this question, we need to determine what these grounding interactions are, apart from being causal events. Must they involve the subject's consciousness of the world? Must they be intentional relations involving perception and action? How does consciousness or intentionality determine intentional relations between the mind and world, and particular relations between thoughts and states of affairs?

In the first two chapters, I will consider and resist two approaches. The first suggests that the grounding interactions are intentional. This approach avoids one hurdle, namely bridging non-intentional relations and intentional relations. The challenge here is to show that there are

intentional grounding interactions. Since these grounding intentional relations are subject-world interactions, by definition these intentional states cannot be empty. If they were, they would be among those which require grounding. Their function requires that they are object-dependent. This conflicts with our intuition. Though many of our thoughts do have referents, all thoughts have the possibility of being empty. Intentional non-existence is a mark of intentionality. Strong arguments for this approach are found in Gareth Evans' *Varieties of References*. These will be the focus of the first chapter.

The second approach denies that the grounding interactions need to be conscious or intentional; it construes them as physical interactions. Intentionality of thoughts is cooked from physical relations. If this can be done, then we will find intentionality directly grounded in causal relations between the subject and the world. I think the most promising materialist theory is Fred Dretske's in *Explaining Behavior*. The second chapter will examine this work.

In the third chapter, I will present my alternative. On my view, the subject-world interactions as grounding interactions need to be conscious (in a non-intentional sense). Intentionality arises from (and hence is grounded in) conscious awareness and behavior. It shows how causal relations and non-intentional consciousness are necessary for the grounding (not merely the existence) of intentional states, but without reducing intentionality to either.

The fourth chapter draws some conclusions from my positive approach regarding reference, sense, the holistic picture of mental content, and rationality. It also tries to show how my theory applies to some contemporary debates: direct vs. indirect perception, conceptualism vs. non-conceptualism, and narrow vs. wide content.

2.0 OBJECT-DEPENDENT THOUGHTS

Evans's concern in *Varieties of Reference* and mine are not the same. He is concerned with showing us how to think about objects in a direct, non-descriptive, manner. He tries to show that there are object-dependent thoughts in perception. If he is right, then these thoughts can function to ground mental content. I will resist the idea of object-dependent thoughts.

The nature of direct reference in perceptual thoughts has the consequence that they are object-dependent. If the referent of a perceptual thought does not exist, then that subject cannot be counted as having a genuine perceptual thought. If Evans is right about the object-dependency of perceptual thoughts, then we have, in Evans's theory, the shape of an answer to our problem. The fundamental object-subject relation that grounds intentionality is the subject's thinking about the object in perception. The grounding relation is already construed as intentional. There is no question of how intentionality arises from a non-intentional grounding relation. The issue left now is how object-independence can be accounted for by the alleged object-dependency of perceptual thoughts.

I will argue that we cannot find a satisfactory solution in Evans's theory. Evans's notion of object-dependency is problematic. But it comes so close to what we need, that it will be informative to examine Evans's arguments for it. In showing how it falls short of what we need, we get a sharper picture of our problem.

In the first section, I lay out the foundation of Evans's thoughts in his discussion of Frege and Russell. This section provides us a motivation for the theoretical need for direct reference. Section 2 examines the general, and relatively abstract, nature of object-dependent thoughts. Section 3 discusses how object-dependency works in perceptual thoughts in particular. In the last section, section 4, I argue that Evans's object-dependency is not well-grounded, and suggest a new direction for our search of the fundamental subject-object relation.

2.1 FREGE-RUSSELL BACKGROUND

2.1 .1 Semantic value

According to Evans, Frege's starting point in *Begriffsschrift* is the thought that the significance of sentences lies in the fact that they can be true or false. An expression is meaningful insofar as it contributes to the truth-value of sentences in which it occurs. This semantic power of expressions is understood to be determined by the expression's association with an extra-linguistic object, which was generally called 'semantic value' (8). According to this account, all meaningful expressions have semantic value. For referring expressions, semantic values are their referents. This theory seems to be committed to the idea that no genuine referring expressions can be empty. A referring expression which fails to refer to any object does not have semantic value. Because possession of semantic power requires having semantic value, such expressions have no semantic power, i.e. make no contribution to the truth-values of sentences they appear in. Suppose someone says in the year 2000 "The King of France is bald". Since France has no king in 2000, the referring expression has no referent, and according to this theory, no semantic power. Consequently the sentence has no truth-value. According to Frege's starting point,

linguistic significance requires truth-value. Therefore the sentence has no linguistic significance. These consequences are highly counter-intuitive. We would want to say that the above sentence and its referring expression are meaningful, regardless of whether the singular term is empty or not. In general, definite descriptions constitute an important class of expressions which can be meaningful even when empty.

Frege first responded by setting such definite description expressions aside for separate treatment. He attempted to treat their meaning as fictional (28). But later he found a more satisfying solution in the distinction between sense and reference. Sense was not initially introduced to deal with the problem of empty terms. It was motivated by the fact that, when confronted with two singular terms with the same referent, a rational person may understand one term but not another. For example, one can know what "Hesperus" means and use it perfectly well in sentences, but not know "Phosphorus", though the two terms refer to the same celestial entity. This cognitive difference is reflected in the fact that for such a person the claim "Hesperus is Phosphorus" is a genuine discovery, rather than a tautology such as "Hesperus is Hesperus." Co-referring terms are problems for Frege, because the linguistic significance of referring expressions tracks their referents. Two terms with the same referent have the same semantic value. But if they have distinct intelligibility, they seem to bear different linguistic significance.

Frege's solution was to elaborate this initial theory of reference to include another tier, a theory of sense which accounts for the cognitive difference between co-referring terms. According to Evans's interpretation, Frege captures this cognitive difference by the idea that a single referent may be presented to a rational subject in different ways. In the case of 'Hesperus" and "Phosphorus", the same planet is presented to the viewer in two different ways, as a body appearing in the morning sky, and as one appearing in the evening sky. Our understanding of

an expression depends on the mode of presentation. Hence one may understand "Hesperus", but not "Phosphorus". On this reading, sense is the mode of presentation (15).

Once Frege introduced sense, he thought he had a new way of dealing with empty terms. In the initial theory of meaning, intelligibility was understood in terms of reference of singular terms. Now we can account for intelligibility in terms of sense. Even though such expressions have no referent, they have sense.

However, Evans argues that this treatment of empty terms is incoherent with the idea that sense is a mode of presentation. Without a referent, there is no mode of presentation of referent, and hence no sense. Such a notion of sense is of no help for empty terms. Empty terms have no sense, and hence are unintelligible.

Furthermore, Evans argues that the idea that expressions can have sense without semantic value is inherently untenable. Semantic value was introduced to aid investigation of the significance of sentences. What counts as a linguistically significant, i.e. meaningful, expression is an intuitive matter. Our intuition says meaningful expressions are those with linguistic viability. In so far as an expression has linguistic viability, it must fall under the class of meaningful expressions covered by our theory (23). According to Frege's theory, the semantic power of expressions lies in its association with their semantic value. Hence it is not open for Frege to claim that some expressions are meaningful (have sense), but have no semantic value (no referent). Frege's semantic theory is committed to the idea that every semantically significant referring expression has a referent. If we want to account for empty terms, we must revise our semantic theory, not just add on a theory of sense.

2.1.2. Russell

It is Russell who provides the requisite revision, according to Evans. On this theory, there is no gap between semantic value and sense. All expressions with sense also have semantic value. Genuine referring expressions have referents as their semantic value. They are object-dependent. If there are no referents, then they have no semantic value, and the sentences which contain them have no meaning. Evans calls these expressions “Russellian” (46). Expressions whose meanings seem to be object-independent, such as definite descriptions, have semantic values, but they are not their referents. To make their semantic value perspicuous, the semantic value of sentences in which they feature needs to be decomposed into the semantic value of sentences which contain genuine referring expressions. So object-dependency is still required for meaning at the foundational level. That connection makes it possible to provide, via Russell’s analysis, object-independent meaning of non-Russellian expressions.

2.1.3 Philosophy of Mind aspect of Russellian terms

For Russell, "Russellian" singular terms are motivated by epistemological considerations. He held that the subject cannot think about an object without knowing which object is in question. Evans calls this claim "Russell's Principle" (65). He takes it to be the key to understanding how we directly refer to objects in thoughts. For Russell, such knowledge is satisfied either by direct acquaintance or by description. The distinction between thinking of an object by acquaintance vs. description is the mental equivalent of the distinction between genuine referring expressions and non-genuine referring expressions. When one thinks about an

object via description, the object can fail to exist without threatening the thought (e.g. "The king of France is bald"). However, if one takes oneself to be entertaining a Russellian thought 'Fa', but 'a' is empty, then the subject is, in fact, not having the thought 'Fa'. Russell was a Cartesian and was committed to mental transparency. He could not make sense of the idea that one can take oneself to be thinking p but not actually be thinking p (44). Russellian thoughts cannot suffer reference failure. Since our thoughts about external objects do not have that kind of immunity, Russell believed that Russellian thoughts are limited to private thoughts, thoughts about sense data.

Evans holds onto Russell's Principle without Russell's Cartesian commitments. He does think that one can take oneself to have the thought p while actually failing to have that thought. So he can admit to the possibility of Russellian thoughts about external objects. Evans takes it to be intuitive that such thoughts exist, paradigmatically in perception, where we have direct acquaintance with the world.

2.1.4 Conclusion

The underlying thought in Evans's reading of Frege and Russell is that reference is primarily an object-dependent relationship. This trivially follows from Frege's starting point that semantic value lies in association between a term and an extra-linguistic object. Considerations of empty terms never challenged the idea that there is a class of terms whose reference is secured in an object-dependent manner.

What Frege and Russell have shown is a theoretical need for a class of object-dependent singular terms. But that does not justify the belief that such a class actually exists. Language is

an everyday phenomenon, a subject matter given to philosophy. If there are Russellian singular terms, then they are part of our everyday language, something to be discovered, not posited. The fact that there is a need for Russellian terms does not show that they exist. What it teaches us is the function of Russellian terms if they exist, and the nature of a philosophical problem, if they do not. Our interest is in philosophy of mind; the same considerations apply here. Russell's *positing* objects of acquaintance at most dresses the theoretical need for them in a different form. It does not say what is true of our mental lives. On this score, Evans is better in looking for object-dependence in our everyday thoughts.

I agree with the motivation laid out here for the theoretical need for Russellian thoughts. I also agree with Evans's approach in justifying the existence of Russellian thoughts by exploring the nature of particular thoughts. However, I do not think such a need is fulfilled. But first, we turn to Evans's claim that there is a class of thoughts which are Russellian, and his arguments for their object-dependency.

2.2 OBJECT-DEPENDENCY

2.2.1 Information-based thought

Sometimes when a subject entertains a thought about an object, she takes her information to stem from an object, unmediated by concepts. Evans calls these information-based thoughts (121). In entertaining an information-based thought, the subject is in an information state, and identifies an object to which she attributes the information. She has an (implicit) conception of what it is for

some information to be derived from an object, and conceives the object as the source of her information. These "implicit conceptions" do not require that subject employ concepts such as "source of my information". Rather they are manifested in the subject's ability to treat the object as the one whose properties determine the truth-values of her thoughts about that object (122). This ability does not require that the subject thinks of the object under any definite descriptions.

Take a simple case. Suppose the subject is looking at a cup and thinks "This cup is dirty." She has a bit of information about an object (call it X). She takes it that there is an object, X, in the world which is a cup and is dirty. She picks out this object in virtue of her ability to conceive and trace an information link to the object of origin. Her mode of identification is non-descriptive. She does not pick out the object as the one which fits the description "X is a dirty cup over there." Rather that piece of information is just the first one the subject obtains from that object. Her conception of X may grow to include more information, such as "It belongs to Jane", "It has been misplaced", etc. None of these pieces of information fixes her object of thought. Which object she picks out is determined by which object she treats as the one which determines the truth-value of her thoughts.

2.2.2 Information

In an information-based thought, the veracity of the information is determined by comparing it with the object it is derived from (which I will refer to as "the source object") (125). The causal link between the object and the information state fixes the object of information without active contribution from the subject. What endows representational, normative status to an information state? Since representation and normativity are inherent to

information, we should ask what makes a state an information state? Very generally a state is an information state in virtue of being an output of the right kind of input-output system. “Now this structure can be discerned whenever we have systems capable of reliably producing states with a content which includes a certain predicative component, or not, according to the state of some object” (125). What makes an input-output system an information system is that it exhibits a structure in which there is a reliable production of outputs which contains predicative element of the input object. Any system exhibiting this structure is an information system.

When an information generating mechanism malfunctions, the output still counts as information in virtue of being the output of such a system. In some cases, the output contains garbled content. In other cases, there may be no input which the output is derived from. In the latter case, the information is of nothing (128).

For mental information states, there is a variety of “mechanism” which provides information links. They play different roles in our communal information system. Perception gathers information. Memory retains information. Communication transmits information. The former is a direct link between a subject at the time of thought and the object of her thought. The latter two provide indirect links between them. All these information links, directly or indirectly, enable the subject to think of an object in a de re, non-descriptive manner.

2.2.3 Thoughts and the Generality Constraint

Mere information link is not sufficient for the subject to think about the object. The content of the information state only provides the predicative component of a thought. It should be characterized as open sentences. The information in the thought “That man is wearing my hat”,

the information content is: “x is wearing my hat.” (124-5). In addition the subject must provide the referential element by picking out an object which she takes the information to be derived from.

These two components of information-based thoughts reflect compositionality of particular thoughts in general. Evans explains this compositionality in terms of joint exercise of two distinct capacities: the capacity to think of the object referred to, and conceptual capacities involved in understanding the predicate (101). To have the thought “John is happy”, the subject must have the ability to think of John, and must know what it is for something to be happy. Part of what it is to possess each capacity is the ability to combine it in an indefinite number of thoughts. For example to be able to think of John, one must be able to not only have thoughts such as “John is happy”, but also “John is strong”, “John is heavily in debt”, etc. Evans refers to such a capacity as “the Idea of an object”: “An Idea of an object, then, is something which makes it possible for a subject to think of an object in a series of indefinitely many thoughts, in each of which he will be thinking of the object in the same way” (104).

The same is true for the predicate component of the thought. To know what it is to be happy is not merely to know what it is for John to be happy, but to know what it is for anyone or anything to be happy. So the subject can entertain thoughts such as “Mary is happy”, “Michael is happy”, etc. In Evans's terms, this capacity is “the Idea of being happy” or “the concept of happiness” (105).

In order to know what it is for Fa to be true, the subject must jointly exercise the Idea-of-a, and the concept F. That knowledge is a fundamental requirement for the ability to think Fa. Hence to have particular thoughts, the subject must exercise a coherent Idea-of-an-object and a coherent concept of the predicate. Evans calls this the Generality Constraint for particular-

thoughts (100). Our focus will be on the Idea-of-an-object. The following section explains what having such an Idea involves.

2.2.4. Idea of an object

Evans's starting point is Russell's principle, which states that a subject cannot make a particular-judgment about something unless he knows which object his judgment is about (89). For Evans, this knowledge is practical knowledge. To be able to pick out a single object, the subject must know how to discriminate that object from other objects. Evans spells out that knowledge in terms of two requirements: the subject must know 1) what kind of object it is, and 2) how to distinguish the object from others of its kind (106).

An object falls under different kinds. A chair is an artifact, a wooden object, a material object, and so on. Each kind is a species of a more general category. These nested categories must terminate at some fundamental kind. For the chair, the fundamental kind is material objects. To fulfill the first requirement the subject must have a conception of the fundamental kind of the object in question. In the case of the chair, for example, she must know what it is to be a material object, and that the chair falls under that kind. In virtue of knowing this, she can discriminate it from objects of other fundamental kinds, e.g. numbers, colors, etc.

To differentiate an object from its fellow objects within its fundamental kind, she knows the individuating property of that kind. For material objects, it is their spatial-temporal locations. Evans calls this individuating property the "fundamental ground of difference". When S thinks of an object in terms of its fundamental ground of difference, S is exercising her fundamental Idea of the object (107). Evans's notion of fundamental Idea helps us understand how the subject

can pick out an object directly, without resorting to a definite description. In exercising a fundamental Idea of our chair, the subject is thinking of it in terms of its spatial-temporal position. Her ability to identify the object's spatial-temporal location is not (fundamentally) manifested in her ability to think "The object has latitude X and longitude Y," but in her disposition to move about the object. So children can know spatial locations of objects and hence can refer to them directly without having concepts such as latitudes and longitudes.

Fundamental ideas are employed implicitly even when the identification is non-fundamental (109). Suppose S is thinking (via definite description) "The President of the United States is tall." She doesn't know the president's location. Let's suppose further that she doesn't know anything about the man. If she were standing right in front of him, without further information, she would not be able to identify him as the object of her thought. Nevertheless, she has discriminating knowledge of the president, because she knows what it is for a unique object to fit that definite description. That requires that she knows how to differentiate a material object from other material objects. So even though she is thinking about the president descriptively, she is still exercising her fundamental Idea of an object, only the mode of identification is not fundamental.

One might think that once the subject has identified an object and predicated the information of that object, the subject has a complete particular-thought. If the mode of identification is of the right sort, then she has a complete information-based thought. However, according to Evans, more is required. For the subject to have a particular-thought, she must have a coherent Idea-of-an-object and a coherent concept. Evans adopts an interpretationist approach to thoughts. So the possession of a thought or an Idea is understood in terms of attribution of

thought or Idea. The main question then is: What does it take to attribute a coherent Idea-of-an-object?

2.2.5 Interpretation and purpose

For Evans, the concept of thought has its home in our ordinary discourse where we make sense of each other's speech (109). Evans's interpreter is not a theorist standing outside the linguistic activities of the speaker. Rather she is a speaker amongst other speakers. Interpretation is just a part of how we understand each other's speech. To understand someone's speech act, we need to understand its purpose. What is the point of that utterance? Consider Evans's example:

A young student is reading out an ill-prepared essay to his class. It contains the sentence 'A spark is produced electrically inside the carburetor'. 'That is not right', the teacher says. 'What does he mean, class?' And here someone may say 'He means the cylinder, sir' (130)

The classmate is not attributing the thought "A spark is produced electrically inside the cylinder" to the student. He is pointing out that this is the thought the student should have, given the student's purpose in his essay (129-30).

An analogous point is true for reference. When a speaker expresses an information-based thought, we may safely assume that she wants to be talking about the object the information is derived from. Therefore, the purpose of her expression is to refer to the source object. Evans thinks her referential success is a 'basis' for speech act. Hence her expression and her attempt at her thought are well-grounded only if the reference is successful (131). A successful reference is one in which the subject picks out the object she is supposed to pick out. If she fails to pick out the right object, either because she picked out the wrong object or no object at all, then she

cannot be credited with the information-based thought she is supposed to have. Because thoughts have purpose, failing to achieve that purpose (on a given occasion) amounts to failing to have a thought. So, all information-based thoughts are object-dependent in this sense: if the subject fails to pick out the source object, then the subject cannot be credited with the putative information-based thought. Depending on the mode of identification, there's another kind of object-dependency.

2.2.6 Object-dependent Idea-of-an-object

We said earlier that having an Idea-of-an-object requires being able to differentiate it from all other objects. Some modes of identification employed to differentiate an object are object-independent. That is, they can deliver a coherent Idea-of-an-object without referential success (136). One case is the use of description in information-based thoughts. Though in all information-based thoughts, the mode of identification is always unmediated by definite description, a subject may still use description to pick out an object without using that description as the determination criterion. Suppose Ted tells Mary to meet Fred at the airport. Mary has never met Fred. But she's told he will be carrying "The Iliad". With that piece of information Mary tries to identify Fred. She is not trying to, nor is she supposed to, pick out the person in the airport carrying "The Iliad." Suppose Fred left his copy of "The Iliad" on the plane, and is wandering about without a book in hand. But there is one (and only one) person who is carrying a copy of "The Iliad", Sam. It is still Fred, not Sam whom Mary is supposed to pick out. But in this case, Mary has been deprived of the means of picking him out. This referential failure does not threaten her Idea-of-Fred. She knows what it is for "Fred is carrying

“the Iliad” to be true. The content of her would-be thought is coherent. The reason why we cannot say this is the content of her thought is because we cannot credit her with the thought on account of failing to pick out the source object.

For some modes of identification, unless the subject picks out the right object, she fails to have the Idea of that object (136). Demonstrative identification is the clearest case. Suppose the subject is hallucinating a cup. She thinks “That cup is dirty”. Recall that Evans's notion of information state is defined by the information-producing mechanism, in our case the mode of identification. Since the mode of identification is such a mechanism, the output is an information state. So the subject is in possession of an information state. But that information state is of nothing. There is no source object to be picked out. But the nature of demonstrative identification is such that we can say if she is hallucinating, then she doesn’t know what she is talking about. In Evans's words, she fails to satisfy Russell’s Principle. She cannot be credited with a coherent Idea of an object, and hence a coherent thought content. She doesn’t know what it is for “That cup is dirty” to be true. So referential failure in this case leads to failure of thought on two accounts 1) failure to fulfill the purpose of the thought, and 2) failure to have a coherent Idea-of-an-object.

2.3 DEMONSTRATIVE THOUGHTS

Our goal is to understand how causal links enable the answerability relation between a subject's thoughts and their objects. For that purpose, our exploration of the object-dependency of information-based thoughts is too abstract to be sufficient for our project. What we have been told thus far is that causal links (in the form of information links) and the right mode of

identification will result in object-dependent thoughts (information-based thoughts). We want to know how (and which) subject-world relation grounds answerability of thoughts. Not all information links can provide that grounding. For instance, some modes of identification function to transfer information (e.g. those exercised in memory and communication). These modes of identification presuppose the availability of information thought. To understand the subject-world connection, we want to know how information link and mode of identification make information-based thoughts available in the first place. To understand that, we must turn to demonstrative identification and demonstrative thoughts in perception. The goal of this section is to understand how the nature of demonstrative identification leads to this kind of object-dependency.

2.3.1 Demonstrative Idea and demonstrative identification

Under typical circumstances, when a subject is situated vis-à-vis an object, perception provides an information link connecting the object and subject. That information link enables the subject to locate the object in space at the time of her perceptual thought. If 'locating an object' just means 'knowing where an object is', then one can locate an object without being situated vis-à-vis the object. Memory, for example, also enables a subject to locate an object at the time of her thought. Suppose A tells B at time t, "Your book bag is where you left it this morning." By remembering where her book bag was earlier that day, B is able to locate that object at time t. What is the significance of being situated vis-à-vis an object?

The significance seems to be that when the subject is so situated, she knows where the object is in virtue of being able to point to the spatial region the object occupies. She cannot

point to a spatial region just by remembering where it is. In other words, demonstrative identification of an object presupposes demonstrative identification of a position in space. The latter requires having an information link with the object which is situated vis-à-vis that object.

I use the idea of "pointing" to help accentuate the directness in demonstrative identification which is absent in identification through memory or description. However the idea is too loose for our explanation. I can point to a chair next to me and think "That chair is broken". There is no difficulty in thinking that I have identified the relevant spatial region. When I look up at the moon, point to it and say "That moon is bright", there seems to be reason to be less confident to think that I have successfully pointed out a spatial region. What is the difference? In the next few sections I will explore the nature of place-identification, which is so crucial to demonstrative identification of objects. Only then can we adequately parse out the idea of 'pointing' in a precise sense.

2.3.2 Place identification

Any adequate Idea of a physical object requires having an Idea of the place it occupies. Because the object is construed as objective, the Idea of a place is the idea of an objective place. So the subject must have an objective conception of space, and realize that when she picks out a place, she conceives it as a position in objective space. When a subject conceives space in an objective manner, each position is represented equally. No perspective is adopted which would make any position privileged. In fact, it is conceived from no perspective at all, not even God's perspective (152). To think of a place objectively requires realizing that it is a position in a perspective-free, objective space, where each position is the same as the others.

When a subject uses demonstrative concepts such as ‘this’ and ‘that’, or ‘here’ and ‘there’, not only is she representing a place as an objective place, but she is also identifying it relative to herself. She knows that what place is picked out by ‘here’ depends on where she is. As she moves, ‘here’ moves with her. Therefore to understand demonstrative concepts, the subject must also have a subjective conception of space, which Evans calls “ego-centric space” (154). Evans's talk of ego-centric space vs. objective space is convenient but can be misleading. Such talk might encourage the reader to think in terms of two kinds of space. But we must bear in mind that this way of talking is really just an expression of two kinds of conception of one (the) space.

When the subject conceives space in an ego-centric way, she conceives the space as centered on her that is she takes herself to be the origin of spatial coordinates. Different places are represented as “above or below me”, “to the left or the right”, etc. (154). Such representations do not require objective representation of space. Suppose a subject is sitting in a pod that is spinning rapidly so that the subject loses all sense of objective space. Inside the pod, she can still point to an apple in front of her and think “This apple is ripe.”, or “I’ll bring it here.” She has no objective conception of where the apple is. She only knows how to locate it with respect to herself. That is, she thinks of the place ego-centrically, but not objectively. Evans refers to thoughts about places in ego-centric terms as “here-thoughts.”

Here-thoughts have the function of coordinating perception and action. It’s fairly clear why action and perception require such a conception of space. Actions radiate outward from the subject to the objects acted on. Perception collects information from objects to the subject. In these contexts, what matters is the subject's relation to objects centered on the subject.

This function is not merely a useful feature of here-thoughts, but part of their essence. The subject's ability to represent a place ego-centrically is reflected in knowing how to act with respect to it. If she hears a sound coming from her left, she knows to turn to the left to look for the source of the sound. If she sees an object above her, she knows where to reach to get it. "[H]aving spatially significant perceptual information consists at least partly in being disposed to do various things" (155). We might say we conceive places ego-centrically through our reaction to sensory inputs. So, what it is to have here-thoughts lies in having the correct behavioral (action) disposition to perceptual inputs. Hence, Evans refers to egocentric space as "the space of the possibilities of one's action" (167).

The ego-centric conception of space is not sufficient for having an Idea of an object. That idea requires knowing what it is for an object to occupy objective space. So the subject must also have an objective conception of space. The significance of the ego-centric conception is that it enables the subject to relate to objective places.

Objective space is not conceived as dependent on the subject. In the objective conception, every position within the space is equal, none more privileged than another. As we have mentioned before, such a conception is not formed from a perspective standing outside the space, e.g. God's perspective. It is from no perspective at all. It may seem hard to think about space from no perspective. That difficulty has roots, I think, in the temptation to think of space as a theatre in which objects live out their lives. As such our relation to space is as some kind of observer, which necessarily views space from some perspective. To have an objective conception of space in Evans's terms is not to first conceive such a theatre stage, and then conceive each spot as equal to others. The conception of a theatre stage has no role in this

conception at all. It is purely about how we construe the positions and objects' relation with each other as equals. It is about having an egalitarian view of places.

Such an objective conception of space is not how we normally think about space. In my empirical thoughts I represent space objectively. I do not take my perspective as a defining point. I do treat myself as one object among others, with no privilege as far as defining the spatial framework is concerned. In that sense I represent each place objectively. However, I do represent the places from my perspective. My conception is not perspectiveless.
So we should differentiate Evans's notion of objective space from this kind of objective conception of space.

However, my ordinary objective conception of space involves Evans's objective space in so far as I treat my position as equal with others. It also includes an ego-centric component. First it is essentially formed from my perspective. Second, in this conception, I know how to relate to other objects through action and perception. So I know how to relate to their places with respect to myself. Consequently, I can refer to these places as 'here' and 'there', and the objects as 'this' and 'that'. I have an ego-centric conception of their places. At the same time, because I see these places objectively, I conceive 'here' and 'there' objectively. The subjective and the objective elements do not exclude each other. Rather, to use Evans's word, they are "mapped" onto each other. Our ordinary conception of space is a result of mapping my objective conception of space onto my ego-centric conception of space, or in Evans's terms a cognitive map.

My Idea of an object requires the ability to identify a place in my cognitive map. Because it involves an ego-centric conception of space, it will be manifested through my behavioral dispositions towards other objects. However, once the objective conception is

mapped on the ego-centric conception, new kinds of behavior become possible. Now I have a different perspective on myself and other objects' behavior. In my ego-centric thoughts, I would regard my movements in terms of how other objects get closer or further away from me. It would make no sense for me to see myself as moving with respect to them, since my position defines the spatial framework. With a cognitive map, I can see myself as moving with respect to other objects, carrying my ego-centric spatial framework with me as I move among other objects. I can recognize that an object remaining still in public space can change its position in my ego-centric space in virtue of my movement towards or away from it. So I can keep track of an object's position as I move about. Conversely, when I want to get closer to an object, I can be said to want the object's position in my ego-centric space to shift towards it. I know I can fulfill this desire by moving towards the object, by changing my position in objective space. The same idea applies to direction. If I were to see an object to my right, I turn my head so the object is straight in front of me. I am again changing the position of the object in my ego-centric space by moving my head.

Understanding our ordinary conception of space in terms of cognitive maps (i.e. in terms of mapping objective space onto ego-centric space) unwraps the relationship between space and subject that enables identification of places. First, this conception makes explicit the subject's practical capacities with respect to other objects. It's her ability to act on other objects that gives substance to the idea of 'space'. Her ability to pick out a place is first and foremost an ability to pick out a place in her ego-centric space. Her ability to treat herself and other objects as equals (and hence to treat places she picked out as objective) is reflected in her recognition that as she moves about, she is carrying her ego-centric space within an objective space, and that she can see other objects as stationary objects which go in and out of her moving ego-centric space. Hence

that our ordinary conception of space enables us to navigate is not merely a useful feature. It embodies the subject's ability to apply her practical capacity within an objective space.

Evans's notions of cognitive map and ego-centric space enable us to think about picking out a place in a much more clear and precise manner than the metaphor of 'pointing'. To illustrate, let's return to the earlier examples of "That chair is broken" and "That moon is bright." There is a significant difference in the subject's ability to pick out spatial regions in these two cases. We cannot adequately explain that difference by merely appealing to the notion of 'pointing' or use of 'that'. Now we can. The difference is that the chair occupies a place within my ego-centric space, and the moon does not. I have practical knowledge of how to relate the chair's location by walking towards it, reaching out for it, etc. My ability to identify that location requires no mediation from concepts. When I am thinking "That chair", I am mapping a position in my ego-centric space onto objective space.

The moon, on the other hand, is not within my space of possible actions the way the chair is. That does not mean that I cannot in principle act on the moon. I can, if properly trained and given the right opportunity, go to the moon. But this action is not a single primitive action within my ego-centric space, but an action stitched from many actions in my ego-centric space, such as walking towards the shuttle, getting out of the shuttle and stepping onto the moon's surface, etc. Through these series of actions, my ego-centric framework shifts, and different places and objects come in and go out of my ego-centric space. Once I am on the moon, the moon would be within my ego-centric space. There, my ability to pick it out demonstratively is as unproblematic as my ability to pick out the chair. However, from my earthly perspective, the Moon is not within my space of action. What the subject can point to demonstratively is defined by her ego-centric space. If an object is not within her ego-centric space, given her spatial-

temporal position, then she cannot point to it in the sense of demonstratively picking it out. Demonstrative identification only applies to objects with which the subject is situated vis-à-vis, because only such objects are located in her ego-centric space.

Now that we know what place identification involves, we can specify what it is to have demonstrative ideas of objects.

2.3.3 Demonstrative Ideas of objects

If a subject can demonstratively identify a place, then she can demonstratively identify an object occupying it. Suppose she points to a book X, which occupies position p, and says "This is mine." Her demonstrative Idea "This" involves two pieces of practical knowledge:

a) What it is for a place she picks out ego-centrically to be an objective place. In Evans's slightly technical terms: for her ego-centric Idea of the place, p, and arbitrary objective place π the subject knows what it is for "$p = \pi$" to be true.

b) What it is for the object she picks out to be the object occupying that place: She must know what it is for "This = the object at π now" to be true (170-1).

Evans often specifies such knowledge in terms of what it is to know that a certain proposition is true. This manner of specification should not lead us into thinking that Evans is ascribing propositional knowledge to the subject. The kind of knowledge that is required in (a) and (b) is practical. To have knowledge (a) is to know how to navigate around the place. To have knowledge (b) is to know how to treat information from that place as germane to one's thoughts about the object.

For the purpose of our exposition, we have treated the place identification independent of object identification. But the two types of demonstrative identification are mutually dependent. To demonstratively identify a place is to treat information from that place as germane to one's thoughts about the place. It requires an information link, which is only possible if there is an object occupying it. Demonstrative identification of both places and objects can only succeed where there is an information link between the object/place and the subject (not what information links the subject takes herself to have). Therefore, the two pieces of practical knowledge (a) and (b), and consequently a coherent demonstrative Idea, also require an information link and the corresponding source object. These Ideas and their corresponding thoughts are Russellian.

2.4 EVANS' FRAMEWORK OF SUBJECT-WORLD INTERACTIONS

If Evans is right, then we have a general framework for how subject-world interactions ground answerability of thoughts. The interactions consist of the subject's perception of objects. The object-dependent relation is not merely construed as causal, but as intentional. The subject has object-dependent demonstrative thoughts about the object in perception. How that accounts for object-independent thoughts remains a large question, but one we may be hopeful.

I will try to resist Evans's claim that demonstrative thoughts are object-dependent. His claim is based on two intuitive considerations about thought: 1) a thought should make sense of the subject's speech act, and 2) to have a thought, the subject must know what it is for its content to be true. (1) Is used to justify his attribution requirement that to have an information-based thought, she must pick out the object which she is supposed to pick out, the source object. (2) In

conjunction with his theory of what it is to have an Idea of an object is used to show that demonstrative thought content is object-dependent. I will argue that these two considerations do not show that hallucinatory thoughts are not thoughts. Therefore, they do not provide support for object-dependency of demonstrative thoughts.

My arguments against Evans do not prove that demonstrative thoughts are definitely object-independent, or that there is no object-dependent intentional basis for thoughts. However, if successful, they will have removed a substantial challenge to our conviction that genuine perceptual thoughts can be empty. As a result, I think, we should be sufficiently discouraged with this attempt at resolving our problem so as to explore other options.

2.4.1 Methodology

First, I would like to say a little about general methodological issues and then turn to Evans's theory. Evans's theory, at first glance, seems to violate common sense. When a theory conflicts with common sense, it is not always clear which one is wrong. The measure here is this. A theory is itself anchored to some pre-theoretical considerations. If not, then there is a real question as to whether the theory and the common sense intuitions are on the same topic. Evans's theoretical apparatuses are ultimately grounded by pre-theoretical considerations. In that case, the question is whether these pre-theoretical considerations cohere with our intuitions or not. If not, then we need to know which are stronger and which need correction. There can be a real debate here, leading to further philosophizing. But the theory in question cannot be brought to bear on this debate; that would be question-begging.

Evans's thoughts are anchored by two *foundational* considerations on what it is for a person to have thought p: 1) thought p can be credited to the person based on the subject's purpose, and 2) the subject knows what it is for p to be true. In Evans's theory, the two converge because of his *theoretical* claims about thought-content.

Thought-content of a demonstrative thought requires the attribution of the relevant Idea of an object and a concept. If no adequate Idea of an object can be attributed, then there is no coherent thought content. There is no such thing as what it is to know that p is true, because there is no p.

Evans's theories result in highly counter-intuitive claims at the foundational level. If Evans is not begging the question, then we should be able to assess claims at the foundational level independent of his theory. The objection cannot be merely that they are counter-intuitive. For, as we have already commented, intuitions can be misleading. The objection, rather, is that the intuitions Evans's theory violates contain fundamental elements of the intentionality of thoughts.

If it turns out these claims are unfounded, then the theory is unfounded. We will examine Evans's verdict on each of the foundational considerations as it applies to hallucination. Failure here is sufficient to suggest that his conclusion that demonstrative thoughts are object-dependent is not well-founded.

2.4.2 Hallucination as thought

Perhaps Evans's theory's most striking consequence is that in hallucination there is no

demonstrative thought; there is only the essaying of a demonstrative thought. Intuitively we have a deep conviction that hallucinatory episodes are thoughts. They share an important feature with other (unproblematic) thoughts: they have an active role in our mental lives. The subject may use them to guide actions, make inferences from them, report them to others as testimony, etc. Some philosophers would cite these functions as definitive of what it is to have a thought. Evans does not explicitly make any such claim. However, in his discussion of linguistic meaning, he voices a similar line of thought, but applied to linguistic meaning. There, he says that the concept of meaningfulness is an intuitive one, not a theoretical one to be defined by philosophers' theories. The intuitive demarcation is drawn by expressions' viability in our linguistic practice. If an expression has currency in our discourse, then it is meaningful. If we apply the same considerations to thought, then the natural thing to say is that what it is to have a thought is an intuitive, non-theoretical issue, and its intuitive demarcation is marked by viability in subjects' or communities' mental lives. Evans makes no such application, and it is not clear why. In any case, to our disappointment, he makes no intuitive judgment of what counts as thought the way he does for meaningfulness. This makes his theory vulnerable to strong convictions on the opposing side. But mere appeal to such convictions is unhelpful. We need to trace where Evans and our intuitions disagree and agree, and where and why they diverge.

Evans gives two reasons why demonstrative thoughts are object-dependent, and hence hallucinatory thought-episodes do not count as thoughts. First, to credit the subject with demonstrative thoughts, the subject must pick out the source object. Since there is no source object, no demonstrative thoughts can be credited to the subject. Second, all thoughts require coherent thought contents. In the case of demonstrative thoughts, failure to pick out the source object means there is no coherent Idea-of-an-object, and hence no coherent thought content.

2.4.3 Thought attribution and purpose

Evans approaches thoughts from the perspective of interpreting speech. This external perspective is helpful for understanding the problem of intentionality. To be clear on what the problems are, we will look at what seems right as well as what seems wrong in Evans's approach.

A mental episode is intentional if it is about something. When a speaker utters "This cup is empty", and we want to know what she is talking about, we appeal to the third person perspective. If the speaker is hallucinating, then from the interpreter's perspective the utterance is about nothing. How can a thought-episode be intentional, if intentionality lies in being about something, and the thought-episodes is about nothing? This is the problem of intentionality. If the problem has any grip on us, it is because, in part, we take the third-person perspective as having a vital role in determining what it is to have a thought. If we leave out the third-person perspective in assessing thought, then there would be no problem of intentionality. On the other hand, it is a genuine problem because we have a strong conviction that there is thought in hallucination. So the aforesaid considerations do not undermine that conviction. For Evans, considerations from the interpreter's perspective do undermine that conviction. They are much stronger claims; they define necessary criteria for a subject to have a thought.

As with all interpretationist theories, the suspicion is that marginalization of the first person perspective will preclude any successful theory of thought. I share that suspicion. I think the significance of the first person, for attribution of thoughts, lies in the 'mental viability' of thoughts which we have already mentioned. For thoughts to play their role in guiding action, making inferences, communication, etc., they must be *meaningful for the subject*. That is the

subject's perspective which seems to me irreducible. Does Evans violate this kind of subjectivity? If so, is he justified in doing so?

In Evans's interpretationism, thought is attributed in the context of making sense of the speaker's speech. A speech act is a purposeful activity. The purpose driving the speech act is itself located within a network of other purposes, driving linguistic or non-linguistic acts. What purposes a speaker has is also a matter of interpretation. So the attribution of thought is part of a much larger project of making sense of the subject's activities (including the speech act in question).

Thus far, the emphasis on the interpreter's perspective has not completely thwarted the subject's perspective. When we speak of the subject's purpose, even if the purpose is attributed, we may need to know how the subject views the world. In everyday discourse, we do attribute background motivations (i.e. purposes) to agents' actions and do not assume that the agent is aware of them. For example, we might say something like "The real reason why Jack wants to become a doctor is for the money, though he'll never admit that to himself." Here we are attributing a motivation, a purpose if you like, to Jack's choice of going into medicine, which is not something Jack explicitly or intentionally adopts or endorses. It is not an issue over which Jack has first person authority. It often takes greater work, in terms of perceptivity and honesty, for the agent to recognize his own deep motivations. Still, the attribution of this deep motivation is based on actions which are essentially performed from the first person perspective, such as deciding to apply to medical school. To understand such actions, we need to appeal to how the subject views the world, and part of that project will involve attributing beliefs, but it is essentially in the context of understanding *the subject's perspective* on the world. In our

everyday discourse, attribution of purpose is not unusual, but it is dependent on first-person actions.

Furthermore, the idea of trying to express a thought but failing to do so is not so objectionable. We do, in everyday discourse, sometimes grasp for words which do not come to us, and an interlocutor may step in to fill in the words. In some of these cases, it is not the case that the subject has a complete thought but could not find the words, but that the subject has an incomplete thought. In some cases, such as the student example, the subject is confused. Confused thoughts and incomplete thoughts are still thoughts. We hold the subject's thoughts to objective standards in evaluating whether they are complete, confused, etc. These episodes are thoughts, not merely because the subject takes them as such, but because they have meaning, "mental viability", for the subject.

For us, the idea of attempting a thought is about trying to capture what is going on with the subject's thought process, not questioning that we are dealing with goings on within the realm of thought. For Evans, the idea is used to determine what belongs to the realm of thought and what does not. Only when the referential purpose is fulfilled, does the episode belong to the realm of thoughts. This move severs any connection interpretation has with thought's meaningfulness to the subject. The fulfillment of the subject's referential purpose is not necessary for the thought-episode to be meaningful for the subject. In making such a fulfillment a requirement for thought, Evans makes meaningfulness per se an inessential element. This is the crux of our disagreement. What is in favor of Evans's view?

Evans's point is that fulfillment of purpose is necessary for making sense of the subject. Since thought attribution is part of that project, fulfillment of purpose is a criterion for thought attribution. But the idea of making sense of the subject cannot support this requirement. Here is

an alternative way of thinking about making sense of the subject which leads to the opposite conclusion.

A subject's speech act is intelligible as long as we can say what she is doing, which includes what she is trying to do. Take again Evans's student example.

> A young student is reading out an ill-prepared essay to his class. It contains the sentence 'A spark is produced electrically inside the carburetor'. 'That is not right', the teacher says. 'What does he mean, class?' And here someone may say 'He means the cylinder, sir' (130)

Given the student's purpose in his essay, the student is trying to convey a fact about how the spark is produced. That is sufficient to make sense of the student's activity. Having a purpose and attempting to fulfill it are sufficient for intelligibility.

The attribution of a mistaken thought itself can be justified by a purpose of the speech act, but a different kind of purpose. As with all actions, speech acts have different levels of purpose. What is the student trying to do? Apart from the one we mention, he may also be trying to write a good paper, to pass his class, etc. There are broader purposes than trying to convey a fact. In the other direction there are also narrower purposes. The most narrow one (or at least close to the most narrow) is his immediate intention. He intended to say, "A spark is produced electrically inside the carburetor", not to say "The spark is produced electrically inside the cylinder." According to this purpose, the subject is supposed to pick out the carburetor, not the cylinder. That goal is satisfied.

If we take this approach to purpose, then instead of saying, as Evans would say, "He should have picked out the cylinder, not the carburetor", we would say "The student should have adopted the 'narrow' intention of saying 'A spark is produced electrically by the cylinder', because he has the "wider" purpose of trying to convey a fact about where a spark is produced." On both accounts the subject did something he should not have done. On Evans's, he picked out

the wrong object, and fails to have a genuine thought. On mine, he adopted the wrong 'narrow' purpose, but picked out the right object. He has a genuine, but mistaken thought. The question is: why should we prefer the 'narrow' purpose approach?

The narrow intentions have greater explanatory power for the subject's actions. The point is better illustrated with a hallucinatory case. If we take S to have the hallucinatory thought "That cup is empty", then we can explain why S's thirst is motivating her reaching out into that region of space. Evans wants to say that S attempts but fails to achieve the thought "That cup is empty." But that does not help explain her action of reaching out her arm. So for the purpose of making sense of the subject's action, of fitting together the subject's various acts into an intelligible whole, we ought to be concerned with the narrow intention behind a speech act, not her wider purposes. If narrow intention guides thought attribution, then referential success would not be a criterion. What the subject intends depends on what objects he takes to be there, not what actually exists.

2.4.4 Coherence of thought content

Another important difference between Evans and me is that he thinks there is nothing to be grasped in any attempt to comprehend what it is for the thought to be true, i.e. the thought's truth-conditions. Truth-conditions are determined by the thought's content. If there is no coherent content, then there are no determined truth-conditions. Hallucinatory thoughts have no coherent content. But it just seems false that we do not know what it is for a hallucinatory episode to be true. Suppose a subject is hallucinating a cup and says "This cup is empty." For the thought-episode to be true, there must be a cup at the designated location and that cup must

be empty. Because no such thing exists, the thought-episode is hallucinatory. Since there are truth-conditions to be grasped, the episode does have content. Why does Evans deny this? Here is the point of divergence. We both agree that having content and having truth-conditions go together. Whereas I approach the issue of content from intuitive considerations about whether there are any truth-conditions to be grasped, Evans approaches the question whether there are any truth-conditions to be grasped from a theoretical approach to content. According to the Generality Constraint, particular thoughts result from joint exercises of an Idea-of-an-object and concepts. Without a coherent Idea to be combined with a concept, there is no coherent thought content. In demonstrative identification, referential failure means the subject cannot have a coherent Idea-of-an-object, consequently no coherent thought content.

2.4.5 Intelligibility of hallucinatory thoughts

What does the subject understand when it grasps the thought “This cup is empty”? How do we specify what the subject knows when we say she knows what it is for “This cup is empty" to be true? There are several candidates, in particular, including the following:

(1) There is a certain object at a certain place that is a cup and is empty.

(2) There is a certain place which is occupied by an object which is a cup and is empty.

If the subject can be said to grasp any of the above, then she knows what it is for “This cup is empty” to be true. (The idea of ‘grasp’ here is not necessarily an intellectual or conceptual matter. The subject can grasp each of these through action.)

There is an important difference and affinity between what it is she must grasp to know “This cup is blue”, and what thought content she actually *forms*, when she is thinking “This cup

is blue". As Evans says, thinking is an *activity*, a joint exercise of an Idea and a concept (in Evans's sense). The referential and predicate components of the resulting thought-episode are fixed by the subjects' choice of Idea and concept. These choices determine what commitments the thought-content makes about the world. In understanding the thought content one grasps these commitments.

As already illustrated by 1-2, commitments come in layers. Here is an important asymmetry between the thought formed and the commitments it carries. In forming her thought, the subject chooses to talk about *one* thing. In doing so, she makes *several* referential commitments. The referential commitments corresponding to (1)-(2) are:

(1a) there is a certain object.

(2a) there is a certain place.

The subject's thought is committed to both there being a certain object and a certain place. It is answerable to how things are with these entities. However, the subject is only *talking about* the object, not the place. If the subject had said "That place is occupied by a cup.", then she would be talking about a particular place. But that is a different thought, involving exercises of different Ideas and concepts. So, the object talked about should not be confused with the entities which the thought is answerable to. This confusion is embodied in the idea of referent as both the object talked about, and the object which determines the normative status of a sentence/thought. In Evans's thoughts, that role is played by the source object. An information-based thought is *about* the source object. At least that is what S is supposed to be talking about. The veracity of the information is determined by how things are with the source object. So the thought content is *answerable* to the source object. The idea of empty thoughts, in particular hallucinations, is that *the object talked about* does not exist. This becomes a crisis if the object talked about is the

object of answerability. If that is true, then the thought is not answerable to anything. Then it would not be a thought at all. But as we have seen that is not the case. The minimum referential success needed for a demonstrative thought to be answerable to the world, is reference to a place. The object does not have to exist. So as long as the subject's exercise of an Idea-of-an-object commits her to a particular place, she has an adequate Idea for a coherent demonstrative thought.

Evans and I take opposite approaches to what it is for a thought to have coherent content. We both agree that a subject has a coherent thought content p only if she knows what it is for p to be true. Evans begins with considerations on what it is to form such a thought, using conceptual tools such as the Generality Constraint, Russell's principle, information link, etc. He argues that in demonstrative thoughts, a coherent content (via a coherent demonstrative Idea) requires the existence of an information link and source object. Therefore, in hallucination, where there is no object, there is no coherent content, and there's nothing that constitutes knowledge of what it is for p to be true.

I, on the other hand, proceed in the opposite direction. I begin with the observation that we do know what constitutes knowledge of what it is for a hallucinatory thought content to be true, and that the thought-episode is answerable to the world. Then we ask what basic requirement does a hallucinatory thought satisfy which enables this answerability. That requirement is the minimum requirement for demonstrative thoughts to be answerable to the world. Any subject's cognitive activity that satisfies that requirement would be sufficient for having a coherent thought. The minimal referential success necessary is reference to a place. Then as long as an exercise of an Idea picks out a particular place, it *must* be adequate for a thought-content which is answerable to the world, or for which there is knowledge of what it is for p to be true.

In claiming that the most fundamental referential relation is between subject and place, we give up Evans's answer to one of our questions: what is the causal or physical relation which grounds intentionality? We began with the conviction that some kind of concrete physical relation between subject and world is necessary for grounding abstract relations such as intentionality. We expect to find this grounding for the most basic referential relation. For Evans that referential link is the unmediated reference to objects in demonstrative thoughts. It is grounded in a causal link embodied in an information link between object and thought. Perhaps, more needs to be said about how these causal links make intentionality possible. That is a further issue. The point here is that Evans's theory has a role for causal relations. Our current picture of demonstrative reference, however, does not. There are no possible causal links between a mere place and a subject. Our sensory apparatus relies on causal links with objects. The referential link between subject and place looks mysterious.

To be clear about the problem, there is nothing odd about the fact that one can refer to a place on a particular occasion without aid from external objects. Suppose my living room is pitch black so that I cannot see anything. I can still navigate it through my knowledge of the furniture in that room, and my ability to represent space. I can think “There’s a chair”, “There’s the coffee table.” etc., without having bumped into these objects, or receiving any type of perceptual input from them. Picking out places without objects is common enough. There is nothing problematic about that fact in of itself. The problem arises when such referential relations play a *grounding* role for intentional thoughts.

Our problem seems sufficient to show that such referential relations cannot be the grounding relation we are looking for. What is the alternative? If a subject has the dispositional *capacity* to refer to places, then she can exercise that capacity on a particular occasion without

input from external objects. That capacity is the capacity to pick out and represent a place as a place. So the subject has an idea of place independent of object. Only then can she comprehend (1) and (2), which requires that the subject be able to represent an object as occupying a place, and represent a place as being occupied (or not occupied) by an object.

Evans has provided us with a good deal of insight into what it is to have the capacity to refer to a place, that is the capacity for place-identification. One needs to develop an ego-centric conception of space and a cognitive map to navigate one's environment. These developments require a network of sensory inputs and behavioral outputs. These are sensory causal inputs from objects, and behavioral causal impacts on objects. So though picking out a place itself does not involve interactions with an object, developing that *capacity* does. Instead of asking how subject-world interactions within particular instances of the subject's representations make these representations possible, we should ask how subject-world interactions in general make possible the subject's *capacity* to represent.

Evans does not explicitly ask the question: what it is for a subject to represent a place *as a place*. Nevertheless his idea of ego-centric space is an important step for answering our question. Ego-centric space demarcates the area where the subject can demonstratively identify a place. Part of such identification is the subject's representation of a place *as a place*. Whereas the objective space is given, the subject's capacity to represent places is acquired. That representational capacity is built out of a network of sensory input and behavioral outputs. So it relies on the subject's physical interactions with objects. The capacity for representing places is developed through a network of interactions with objects. The question is how these interactions enable representational capacities. To answer this question, these interactions cannot be conceived in intentional terms of perception and action. Perception and actions involve exercises

of representational capacities, and presuppose the subject's possession of these capacities. Hence they cannot explain how the subject acquires these capacities. The requisite subject-world interactions must be construed in non-intentional terms. In the subsequent chapters, we will examine efforts to understand intentionality in terms of physical relations.

First, let me clarify the difference between this turn to physical relations and our starting assumption that physical relations underlie intentionality. We began with the assumption that abstract intentional relations must be grounded in concrete object-dependent relations, and that these relations must be built on an infrastructure of causal or physical relations. But we did not require that the grounding object-dependent relations must be *construed* as causal relations, only that they must *involve* causal relations. So the object-dependent intentional relations of Evans's demonstrative thoughts qualify as a candidate, since they do involve causal relations in the form of information links. The fact that these relations are intentional does not in and of itself rule them out as explanans. Our goal is not to explain mental intentionality in general, but object-independence of intentional thoughts. There is nothing objectionable about object-dependent intentional relations explaining object-independent ones. But now that we are looking to intentional *capacities* to explain object-independence, any intentional states or actions (including those which embody causal relations) are ruled out because they all presupposes the capacities we aim to explain. So in order for the grounding object-dependent relations between subject and object to explain object-independence of thoughts, they must be stripped of their intentional layers. The proposal now is that not only do our grounding subject-object relations *contain* causal relations, but that they ought to be *construed* as causal relations to explain intentional non-existence.

What explains intentional non-existence is that the capacities to represent entities as a place, as a physical object, or as anything else can be exercised in demonstrative thoughts without the existence of the object talked about. They are also equally exercised in non-demonstrative thoughts. They are not special tools for demonstrative thoughts but for thoughts in general, and they account for answerability of thoughts in general. Unlike what we had thought, perceptual thoughts' immediacy in representation, directness in reference, and function as our source of information do not entail that their answerability is foundational to that of other thoughts. The closeness between subject and object in perception is special, but not by providing information, rather by enabling capacity to represent. As such, they must be 1) construed in non-intentional terms, and 2) understood as part of a network involving behavioral impacts on objects.

In the next chapter, we will consider a theory which satisfies (1) and (2): Dretske's materialism. According to his theory, the grounding interactions are physical. Intentionality is grounded by them, or arises from them, in virtue of a network of physical counterparts beliefs, desires, and actions.

3.0 DRETSKE'S MATERIALISM

In *Explaining Behavior*, Dretske argues for the indispensability of mental representation within a materialist framework. The explanatory power of mental representation is shown in intentional explanations of actions. Clyde went to the kitchen because he wanted a beer and believes there is one in the refrigerator (51). Such explanations tell us why the subject did what he did by appealing to the subject's beliefs and desires. They are seemingly at odds with powerful scientific explanations of behavior. Science tells us that the real explanation of the behavior lies in a neurophysiological story featuring neurons' firings, muscle contractions, electric impulses, etc. This story seems to have no room for intentional states such as beliefs and desires. From the scientific stance, intentional states seem causally irrelevant.

Dretske nicely illustrates the point in an example. A soprano's high-pitched sound shatters some glass. The sounds she uttered have meaning, but the meaning was causally irrelevant. What caused the glass's shattering was the vibration from the sounds (79). Let us stretch the example with a little imagination, and suppose that the words uttered had the effect of commanding the glass to break. Then, the meaning begins to seem relevant. Some might think the singer caused the glass to break by commanding it. But that is sheer superstition. The real explanation lies in the physics.

Is intentional explanation also a form of superstition? Perhaps our behaviors, like the glass's shattering, are really to be explained by the underlying physical events (in the broad

sense). To think that belief and desire are responsible for our behavior is akin to thinking that the singer's command led to the glass's shattering. In the soprano case, the idleness of meaning is not threatening because we presume it is causally effective in other ways, for example moving the audience to tears. When we question the legitimacy of intentional explanations in general, we are in danger of claiming that there is, in principle, no way for meaning to be causally effective. Most of us want to avoid such conclusions. Dretske tries to show us how.

3.1 THE NATURE OF BEHAVIOR

3.1.1 Movement vs. behavior

Dretske argues that intentional and scientific explanations do not compete with one another. They seem to compete because they seem to explain the same phenomenon: our behavior. But they do not. The behavior science explains is not the behavior belief and desire cause. The latter is action, such as Clyde's going to the kitchen; Sara's playing the piano, etc. These behaviors have a physical component. When Clyde goes to the kitchen, his legs move. When Sara plays her piano, her fingers move across the keyboard. The physical component is *movement*, one aspect of *behavior*. All behaviors involve movements. But not all movements are elements of behaviors. Science explains movement, not behavior. Belief and desire explain a certain kind of behavior, not mere movement.

What distinguishes movements that are not behaviors from those that are? Some movements are products of the subject's *doings*, and others are events which *happen* to the

subject. When Fred's hand is moved by someone else, the hand's movement happens to him. When Fred moves his own hand, then it is something he does. Behaviors are movements which the subject *does*. According to Dretske, the difference between behaviors and non-behavior movements lies in the location of their initial causes, the events which set off the chain of causal events ending with the movement[2]. The initial cause of non-behavior movements is external to the subject. If Jones moves Fred's hand, Jones is the external cause. On the other hand, a movement which is performed by the subject has an internal cause (2). When Fred moves his own hand, the causes are his mental states inside his head. "*If* we have a well-defined ordinary notion of behavior … it is, with a few refinements, equivalent to internally produced movement or change" (3). Dretske's definition of behavior is quasi-technical. Whether it is in the end the right analysis of our ordinary notion of behavior, we need to distinguish it from the latter. When using Dretske's notion, I will either speak of it explicitly as such "Dretske's notion of behavior", or, to be less cumbersome, use "behavior".

"Behaviors" are doings. We are not the only ones who do things. The contrast between happenings and doings also applies to animals, organs, plants, even instruments[3] (94). When a healthy heart pumps blood, this is something the heart does. When the heart's contraction is driven by a pace maker, then it is an event that happens to the heart. When a flower blooms, the blossoming is the flower's "behavior". If it sheds its petals because of the wind, it is an event that happens to it. So, "behavior" needs to be distinguished from human action, and "behavior" explanation from intentional explanations. The latter is a species of the former.

---

[2] The precise identification of the initial cause requires further investigation. But for our purpose, Dretske thinks we can rely on our intuitions.

[3] Dretske's idea of 'what one does' is drawn not just from the way we ordinarily talk, but also biologists, psychologists, endocrinologists, etc. (6)

3.1.2 Behavior as process

The difference between "behavior" and movement is helpful in deepening the distinction between scientific and intentional explanations. However, it is not sufficient to reduce their competition. If behavior (without the star) is a kind of movement, it is receptive to both scientific and behavior explanations. Science explains behavior *as a movement*, and behavior explanations explain it *as behavior*. As a materialist, Dretske needs to show that in explaining behavior *as behavior* we are providing physical explanations. As a supporter of genuine intentionality, he needs to show that these explanations are distinguishable from scientific ones. Thus far, the concept of "behavior" is too close to that of movement to make these distinctions. Dretske needs a wider gap. He shifts to another conception of "behavior"[4]. This second conception is used in the rest of the book.

On this new conception, "behaviors" are restricted to internally produced movements. However, they are not these movements themselves, but the *process* in which the internal cause produces these movements. When Clyde walks to the kitchen, his mental states set off a chain of events, which ultimately leads his legs and feet to move towards the kitchen. Movement is the final event in this process. "Behavior" is the entire process (29), beginning with his mental states and ending with his leg movements. Dretske schematizes the process as C → M, where C stands for the internal causes, M stands for the movement, and → stands for causal process C → M (17).

Why should we identify behavior with C → M, rather than M? Dretske argues that C→M explains better actions in which there is a lag time between what the agent does and the

[4] There is no explicit acknowledgment of this shift. This muddies his view on behavior. But since we are treating his notion of behavior as a technical term, we can set aside his understanding of behavior.

consequence of that action. Suppose an assassin shoots a victim on Oct. 12th, and the victim dies from the gun wound on Oct.15th. The assassin killed the victim. When did this happen? There is a problem here. We cannot say it is Oct. 12th, when the shot was fired, because the assassin cannot be said to have killed the victim, until the victim is dead. The victim did not die on Oct. 12th, but on Oct. 15th. But it also does not seem right to say that the assassin killed him on Oct. 15th. In killing a victim, the assassin *did* something to lead to the victim's death. He did not do anything on that day that led to his death. The culpable action was performed on Oct. 12th. If we are to maintain that killing involves both the death of the victim, and the action which caused the death, then killing must be the *process* which began with the shooting and ended with the victim's death. Analyses of such cases show that behavior is not the cause or the effect, but the causal process (21-2).

This argument is not so convincing since it is based only on telescoped actions. Such actions are built out of events. But what about simpler behaviors such as waving my hand? Pre-theoretically, the physical component is a single event, the movement of my hand, not a causal process linking two events. Dretske would deny this, and claim that there are two events, the initial event happening in the head which leads to the movement in the hand. As a counter-argument this would be question-begging. That claim is part of his theoretical analysis of my waving my hand. The question here is what backs up this analysis? Our intuitions work against it. The only argument he gave does not apply. He has not provided sufficient reason to think that all behavior should be understood as having the form C → M.

3.1.3 Two kinds of causes

For now, let us accept Dretske's claim that behavior is not a kind of movement, but the process C → M. The structural difference between C → M and M shows that intentional and scientific explanations do not compete with one another. They explain different kinds of phenomena. Science asks what causes M? What is the initial event that triggered the causal chain which ended with the movement of my hand? Scientific explanations are concerned with identification of what Dretske calls triggering causes (42), i.e. C.

According to Dretske, behavior explanation, including intentional ones, explains behavior (C → M), not merely movement (M). Suppose someone asks "Why did Clyde go to the kitchen?" ---"Because he has a certain belief and desire that caused him to go to the kitchen." --- "Yes. But why did that belief and desire cause Clyde to go to the kitchen?" The first question "Why did Clyde go to the kitchen?" strictly speaking asks about a movement, Clyde's going to the kitchen. The answer provides the cause C (a certain belief, B and desire, D). That is clearly not sufficient to answer the question. So the interrogator rephrases the question. The revised, real, version asks not why M happened, but why C caused M? Intentional explanations provide what Dretske calls "structuring causes" of behavior.

Dretske thinks that the mistake behind mental epiphenomalism is the assumption that intentional explanations seek triggering causes. In this mistaken line of thought, both mental states (beliefs and desires) and neurological events are taken to provide the ultimate explanation of behavior. There is one kind of ultimate explanation, those which identify triggering causes (115). Science shows us that the real triggering causes lie in the physical properties of mental states. Representational properties now seem causally irrelevant (36).

However, if Dretske is right, there is not just one kind of ultimate explanation, but two. When the scientist asks, “Why did Clyde go to the kitchen?” he is really asking “Why did Clyde’s legs and feet move in such and such manner?” When his friend asks (in intentional discourse), “Why did Clyde go to the kitchen?” he is asking “Why did Clyde’s belief and desires cause his legs to move in that manner?” These are clearly two different kinds of inquiries. The scientific answer, however detailed, does not help address the intentional question. It is the job of representational content to explain why "behavior" (C →M) happens. So scientific explanations cannot take over intentional explanations.

Thus far, Dretske has taken two important steps. First, he has shown how intentional explanations are not under threat from science. Second, he has provided a sketch of human behavior which, though not yet thoroughly natural, is primed for reduction. The skeletal structure C → M uses only natural ingredients: the internal states which constitute C, the physical movement M, and the causal relation between them. His job now is to add representational content in materialist terms, and to show it can be relevant to explaining "behavior".

3.2 REPRESENTATION AND FUNCTION

3.2.1 Indication

Dretske uses indication as the materialist basis of his theory. For an organ to be able to represent the color red, its behavior must track the absence or the presence of redness. This capacity

distinguishes our eyes from our ears, and from eyes of animals which cannot detect color. In general if A represents B, then A must exhibit a non-coincidental covariance or correlation with B. In Dretske's terms, A must indicate B (57).

Dretske's notion of natural indication is akin to Grice's notion of natural meaning. "Grice distinguished what he called a natural sense from a non-natural sense of the word 'meaning.' The natural sense of 'meaning' is virtually identical to that of 'indicate' and that is how I shall normally use the word" (55). Indication as natural meaning is operative in 'natural signs', such as tracks in snow, finger-prints, tree-rings, and cloud formations (54). It is also at work in instruments like thermostats, clocks, and gas gauges, whose engineering exploits natural indication to suit our purposes.

Indication entails correlation or co-variation. In most cases, such correlations are causal and lawful, such as the relation between tree age and the number of tree rings. But they need not be lawful. For example, the ringing of a doorbell indicates that some person is at the door. There is no natural law that governs who rings doorbells. What is important is that the co-occurrence between the indicator and the indicated is no coincidence. There must be some condition which explains this correlation (57).

Correlation is clearly an objective concept. But, in ordinary discourse, the word "indicate" can be used in a subjective manner. The muddy boots indicate to Mary that her husband has been working in their garden. They bear no such indication to her friend Jill because Jill knows nothing about her husband's gardening. In such uses of 'indicate', X indicates Y *to someone*, one who, in some sense, knows that X correlates with Y. For our purpose, we need not argue whether the ordinary notion of indication is subjective or objective.

We can acknowledge that there are two senses of 'indicate'. The question is which one is Dretske using?

Dretske's naturalism requires that he use the objective sense. The subjective sense already imports something materialists have trouble explaining, the first person perspective. So Dretske needs his notion of indication to be objective. He does not provide much discussion on this issue. His only comment is: "[I]f … the registration of this gauge does not indicate what the boiler pressure is to anyone---it nonetheless still indicates what the boiler pressure is" (55).

This claim is quite modest. It only denies that the indication relation of a *particular* indicator requires the appreciation by at least one observer of that indicator. It leaves open the possibility that the indication relation is set up by a user and a class of those indicators. Jack knows that tree rings indicate tree age. We may say tree rings, in general, indicate tree age to Jack. Even for a particular tree ring which Jack has never seen, it still indicates tree age to Jack in virtue of Jack's general knowledge. But if no one has ever seen tree rings, then tree rings do not indicate tree age. Indication is indication to someone. Dretske has not ruled out the idea that indication as indication to someone.

Indication, like causation, suffers from the above two problems for materialists. First, it lacks a normative component. A set of tracks either indicates or does not indicate a rabbit's passing by. Suppose the tracks were constructed by a person, it is not right to say that they still indicate a rabbit's passing by, only in this case the indication is false. The tracks do not indicate at all. There is no *mis*indication, no counterpart *mis*representation (66).

Secondly, for any representational state, the state indicates much more than what it represents. For a particular indicator, there are usually multiple factors that co-vary with it. Dretske illustrates with a gas tank example:

> Electrically operated fuel gauges indicate not only the amount of fuel left in the tank but also the amount of electrical current flowing in the wires connecting the gauge to the tank, the amount of torque on the armature to which the pointer is affixed, and the magnitude of the magnetic field surrounding this armature. Given the way these gauges operate, they cannot indicate ... the amount of fuel in the tank without indicating ... these related conditions (59).

The gas gauge does not represent all these things. It only represents one thing: the amount of gas in the tank.

3.2.2 Function to indicate

Dretske's solution to both problems is to employ the idea of function. The gauge represents the fuel levels, not the other indicated objects, because it has the *function* or the *job* only of indicating fuel levels. Though it indicates many other conditions, it does not function to indicate current flow or torque degree, etc. The idea of 'function' weeds out objects of indication that are not objects of representation.

Function also imports a normative element. If an instrument has a function, then it is *supposed* to behave in a certain way. If it does not behave in the prescribed manner, then it *mal*functions. It seems we have a counterpart to misrepresentation. The question, then, is: What gives a state or an object the function to indicate a certain condition?

An indicator acquires a function when it acquires a job in a system's behavior production. For example, if we take a set of bimetallic strips out of a thermostat, it still indicates, co-varies with its environmental temperature. But we can no longer say it *functions* to indicate the

temperature. Its function relies on its causal role in the thermostat's behavior. That context gives the set of bimetallic strips a job to do. It is given this job because of its indication property. As a temperature indicator, it enables the thermostat to turn the furnace on or off according to temperature change. When an indicator is incorporated into a system's behavior production mechanism, its indication acquires a function, a job. The job is to enable the proper performance of the host system.

In general, an indicator of F acquires the function to indicate F when it is given the job in causing a system's behavior in virtue of its capacity to indicate F. "Once [the indicator] C is recruited as a cause of [behavior] M -- and recruited as a cause of M because of what it indicates about [fact] F -- C acquires, thereby, the function of indicating F. Hence, C comes to represent F" (84). An indicator of F needs to be "recruited" for a job in order to have the function to indicate F. The recruitment occurs when the right causal relation is established between the indicator and the behavior production mechanism. In the case of instruments, that causal relation is set up by the creator, so it is the creator that 'recruits' the bimetallic strip for the job of determining the temperature.

The representational properties of fuel gauges or bimetallic strips depend on external representational systems, namely us, their users and the creators. These objects can represent fuel levels or temperatures only because *we* can represent them. Their representational properties are derived from ours. In Dretske's terms, that means their functions to indicate are assigned by us. To find the 'source' of intentionality, we have to look to original representational systems (67). That is, systems whose functions to indicate are not assigned by other representational systems. We want an account of non-derivative establishment of functions to indicate.

3.2.3 Non-derived representational systems

To understand non-derivative representation, Dretske begins by looking at tropistic organisms, such as frogs or flies. Their behaviors are purposeful, and their biological parts have functions. Unlike intentional behavior, they respond reflexively to environmental stimuli. For example, the noctuid moth's auditory system is designed to detect bursts of high-frequency sounds emitted by bats, their primary predator. When it detects such sounds, it instinctively flies in the opposite direction. The purpose of this behavior is to enhance the moth's survival. The internal indicator, which detects high-frequency sounds, has the job of informing the moth of the presence of such danger. Since the indicator plays a role in the moth's behavior, it not only indicates high-frequency sounds but has the *function* to indicate them, and therefore *represents* the sounds (63).

What "recruits" the indicator into the moth's flight behavior mechanism? At first, Dretske considers natural selection as a possible answer (92). Evolution made sure that moths have these detectors because of their usefulness to their survival. Natural selection is responsible for the fact that the moth's detectors are incorporated into its production of flight mechanism. It is the recruiter.

Dretske rejects this proposal, however. It provides the wrong kind of explanation. Natural selection explains why certain (shared) behaviors, and not others, are what *we find* in nature. They do not, and are not supposed to, explain why *an individual* behaves the way it does. Dretske nicely captures the difference between evolutionary explanations and behavior explanations with the following illustration:

> It is, in effect, the difference between explaining why (all) my friends imbibe martinis, an explanation that requires my telling you something about them, and explaining why I

> have (only) martini imbibers as friends, an explanation that requires my telling you something about me (93).

In behavior explanation, we want to know why my friend, Rob, drinks martinis. The fact that he drinks martinis may be one reason why Rob is my friend. But how I choose my friends is a separate issue. Likewise, why some moth's behavior was 'selected' to survive does not explain why the moth behaves like that in the first place. The right kind of explanation appeals to facts about the individual organisms, not a class of organisms (93). For Dretske, these facts are facts about how the individual develops the behavior in question.

3.2.4 Development

A development may be an acquisition of a behavioral pattern. Development sets up the causal route between the indicator and the system's behavior production. In biological development that causal route is established by biological pathways which may be established prenatal or postnatal. If the behavior is tropistic, a series of events in the organism's history may condition the organism to a certain reflexive behavior. For our purpose, we are not concerned with the specifics of how conditioning occurs. It is sufficient that we have a solution to what hooks up the indication relation with behavior production.

The moth's internal indicators, though representational, are not beliefs. Though the indicators are causally relevant, their contents are not. Beliefs are internal indicators whose representational *contents* are causally relevant to behavior production. We can think of an indicator for F as a switch that flips on or off depending on its detection of F. A switch's causal role lies in its on/off status, which is separable from what leads to that status. For example, a

light is turned on by its switch. Whether the switch is turned on intentionally by a person or accidentally by an object is irrelevant to the switch's causal role. What matters for turning a light bulb on or off is just the switch's position, not what caused it.

It seems we can make a similar distinction between an F-indicator's causal role in terms of its being on or off, and whether it detects F or not. The indicator's effectiveness depends only on its on/off position. The fact that it is turned on by detecting F is irrelevant. The property of indicating F is relevant only to the *recruitment* of the indicator, not to its causal work on a particular occasion, just as Sara's college degree is relevant to her getting her job, but does not determine how she performs her daily work. An indicator has the representational content F, if it detects F on that occasion. Since the detection is irrelevant to the indicator's causal efficacy, so is the representational content. The moth does not fly away as a *response* to that message. It is not even capable of registering it. Its behavior is an automatic consequence of the detector's being 'on'. So even though these detectors have content, the content is not at work[5] (94).

For representational content to do work, what the state *says* must be causally relevant. The system behaved the way it did because of what its indicator *said* (94), because of the *meaning* of the representational state.

> "To qualify as a belief it is not enough to be an internal representation (a map) that is among the causes of output, something that helps us steer. The fact that it is a map, the fact that it says something about external conditions, must be relevantly engaged in the way it steers us through these conditions. What is required, in addition … is that the structure's indicator properties figure in the explanation of its causal properties, that what

[5] There is a problem with idle representational content. According to Dretske what makes representational contents non-epiphenomenal is that they are causally relevant to behavior. That rules out the possibility of causally impotent representational contents. So how can Dretske justify positing representational content where such content is causally irrelevant?

it says (about external affairs) helps to explain what it does (in the production of output)" (94).

3.3 MEANING

How can meaning be explanatorily relevant? Dretske does not reify meaning. The meaning of an internal indicator is causally efficacious, if *having that meaning* makes a difference to how the indicator causes the behavior. "In exploring the possibility of a causal role for meaning one is exploring the possibility, not of meaning itself being a cause, but of a thing's having meaning being a cause or of the fact that something as meaning being a causally relevant fact about the thing" (80).

Dretske's use of the word 'meaning' is a significant and potentially a problematic terminological shift. I had argued earlier that Dretske needs to use an objective sense of 'indication' to avoid importing the first person perspective. We can grant him that partly because 'indication' is ambiguous. However, 'meaning' clearly has a subjective element. The kind of meaning he has in mind involves being *meaningful to* the system/subject it belongs to (95). Now Dretske can no longer be allowed the objective sense of 'indicate'. For him, meaning is what an indicator indicates. That content is meaningful to the subject means that what an indicator says it *says to* the subject. This subjective element requires reduction. However, Dretske does not even seem to realize this problem. This issue looms over Dretske's entire project. We will return to this to explore its connection with other problems in his naturalization agenda. For the purpose of understanding his naturalist agenda, we can set aside the subjective notions of meaning and indication, and think of meaning as what an indicator indicates in the objective sense.

3.3.1 Purposeful behavior

An indicator's meaning is causally relevant only for genuinely purposeful behavior. Tropistic behavior seems to have a goal, but the animal does not aim at the goal. It merely reacts reflexively. If we speak of a moth having purpose, but the purpose is in the eye of the observer. Believers' behaviors are directed towards a goal. Dretske calls such genuinely purposeful behaviors 'goal-directed' (111).

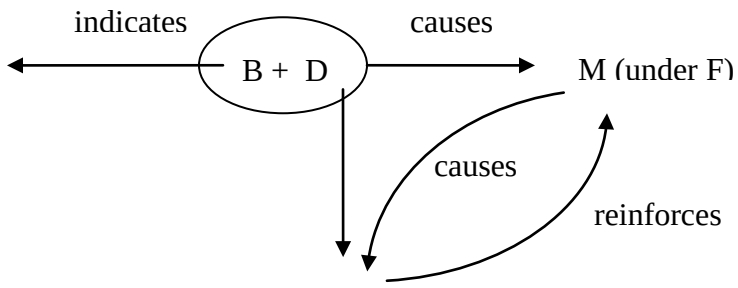
Goal-directed behaviors are (partly) governed by their consequences. "Let us say … that goal-directed behavior is not only behavior that tends to have a certain result but behavior that occurs *because* it tends to have this result" (111). This 'because' is clearly not the 'because' of evolutionary explanation. There is an evolutionary sense in which the moth's flight occurs 'because' of its benefits for the moth. However, there is an important sense in which it is insensitive to its benefits. A blowfly is instinctively drawn to sugar water. Even if, upon one or more occasions, it gets sick from consuming sugar water, it will continue to do so. The negative feedback has no causal effect on the organism's behavior production system (125). The causal connection between sensory stimuli and behavioral responses, once established, is insensitive to the benefits or harm it brings to the subject. Tropistic behavior is blind to its consequences. In that sense it does not occur because of them.

Some animals' behaviors 'reward' or 'punish' the subject. If the result is a punishment for the animal, then it will avoid that behavior in the future. Suppose a blue jay gets sick after consuming a monarch butterfly. It will then avoid eating them in the future, even when it is

hungry. When the subject finds the consequence of its behavior rewarding, the behavior under the same circumstance will occur in increased frequency (125). For example, once a blue jay eats an acorn and finds that it satiates its hunger, it will eat acorns more often when it is hungry. The frequencies of these behaviors are modifiable by the subject's experiences. In ordinary discourse we would say the blue jay learned to avoid monarch butterflies or learned to eat acorns. So Dretske calls the process in which a behavior is modified by its consequence 'learning'. A learned "behavior" occurs because of its consequences for the subject. This kind of modifiability and learning is the defining feature of goal-directed behavior.

For a behavior to be modifiable by its consequences, its causes must be sensitive to these consequences. Two kinds of sensitivity are necessary. First, the cause must include a structure which registers the circumstances under which the consequences of such behavior occur. If a subject's movement M, is rewarding under circumstances F (e.g. eating is rewarding when what is consumed is an acorn), then the subject must have an F-indicator, which will inform the subject of any future circumstances when M-ing may be rewarding. The subject must also possess a structure that recognizes the consequences as rewarding. Call such consequences 'R'. That structure is receptive to R. The receptor will turn on or off depending on the registering of the 'reward' or the 'punishment' (125). The F-indicator and the R-receptor are the physical structures of belief and desire respectively. So Dretske calls them B and D. We can map out the elements of goal-directed behavior in the following diagram:

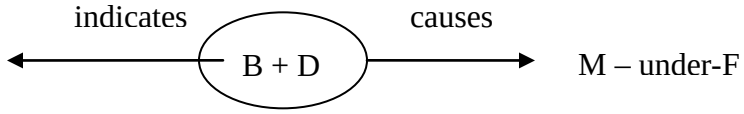
Diagram 1. Dretske's Structure of Learning 1.



The role of B is to inform the subject of the presence of F. If a blue jay wants to eat acorns, then it must know whether there are acorns in its environment. Its behavior will depend on whether the acorn-indicator is ‘on’ or ‘off’. B has a job to do in the production of M-under-F. Therefore, B has the function to indicate F, and counts as representing F.

That much is true for an indicator of a tropistic behavior. We can draw a corresponding diagram:

Diagram 2: Dretske's Structure for Tropisitc Behavior.



In both cases, indicator B is part of M's causal mechanism because it indicates F, and by Dretske's definition, it counts as a representational state. But in a believer, B's *content* is also causally relevant. Her M is controlled by an additional switch, D. If a blue jay sees acorns, but is not hungry, it will not eat. In a reflexive "behavior", the organism automatically M's upon registering F. So any consequence of M is irrelevant and so is any receptor for it. Why should the causal relevance of D and R affect the causal relevance of B's content?

3.3.2 Learning as recruitment

The relevance of D and R reflects a difference in how B is recruited into the effector mechanism. In both types of "behavior", B is recruited because it indicates F. But in tropistic behavior the recruitment does not affect the *way* in which B plays its causal role. Remember B is just a switch, which determines behavior by its on/off position. That connection is set up by genetic, biochemical pathways. Indication properties are not relevant to the set up. Whenever a blowfly detects sugar water, it will automatically drink it. The behavior's causal mechanism is insensitive to such consequences because it is a matter of biochemical hardwiring. Whether one of its structures indicates anything is irrelevant to how that structure is causally linked to the movement.

However the situation is different with believers. B is causally set up with M through learning, not through biological pathways. Prior to learning, the believer has both B and D, and occasionally M. But these components are not connected. If Clyde does not know that baklava is sweet and can satisfy his sweet tooth, seeing a baklava would have no effect on his behavior, even when he is craving for something sweet. His seeing the baklava and his desire for

something sweet are not linked into an effector mechanism. Suppose, upon a whim he eats a piece of baklava for the first time. It just happens that he has a sweet tooth at the time. But his movement is not be caused by the joint power of that desire and seeing the baklava. The subsequent satisfaction of the desire is accidental. Though his behavior requires seeing the baklava, there is, yet, no correlation between his seeing the baklava and eating it. After that initial satisfaction, Clyde learns that baklava can satisfy his sweet tooth. Dretske call this new knowledge "background belief" (116). Such knowledge is implicit and practical, and finds expression in goal-directed activities. This practical knowledge is vital for the building of the effector mechanism. It is the 'glue', as it were, that binds the relevant indication and receptivity. Now when he sees a baklava and has a desire for sweets, he will eat it. His eating a baklava is purposeful, directed toward the goal of satisfying his sweet tooth. His seeing a baklava acquires a new function: to enable him to satisfy his sweet tooth through eating one.

Learning enables a subject to acquire a new "behavior": (B+D) → M. The individual components, B, D, and M are not new. The new elements are their causal connections. The components are now linked to form a new goal-directed behavior. This happens partly because B indicates F (99). B's indication property enables B to acquire its new causal role, confers representational status on B. That B indicates F now means that B has the representational content or meaning of F.

3.3.3 Conclusion

Dretske's tasks are to show how an internal indicator can count as having meaning and how meaning on this view is causally efficacious in intentional explanations. Learning accomplishes

these two main goals. It gives B a role in causing M, thereby turning B's indication of F into B's *representation* of F. Now what B indicates is its representational content and meaning. Furthermore, learning shows how what B indicates is explanatorily relevant to the behavior C → M (under F). Dretske had argued that intentional explanations provide structural causes for behavior C→M, differentiating them from scientific explanations which provide triggering causes for M. In learning, B's indication property is structural, not triggering. What triggers M is whether B and D are 'on'. What B's indication F explains is why B, as opposed to another internal indicator, has the job of causing M-under-F. The reason is that B indicates F. If it had indicated a different fact, say G, it would not have been recruited as part of the effector mechanism.

In differentiating the triggering power of B from the structural cause of B-indicating-F, Dretske is treating belief as involving two components: a vehicle carrying content. These two components are distinct in the sense that they have different, though obviously related, causal powers. The vehicle as an internal indicator provides a triggering cause, and can do so even when the content is causally idle. In believers, the content provides a structuring cause, as an addition to the vehicle's triggering cause. In differentiating the causal power of a belief's vehicle from that of its content, Dretske is able to draw the distinction between structuring causes and triggering causes, and hence between intentional and scientific explanations.

Dretske's naturalization is interesting and significant because his idea of mental intentionality is robust and undiluted. According to his account, instruments such as thermostats and lower level animals such as frogs are not intentional systems. If one is willing to include such systems within the mental, then not only is one's concept of the mental counter-intuitive, but, more importantly, it stretches the concept beyond recognition and beyond our subject of

interest, *our* mental lives. To be sure, there is no clear intuitive boundary between organisms which count as having a mind like ours, and those who do not. But the vagueness does not threaten the distinction between mental systems and non-mental systems. To keep the subject of our inquiry on track, we need to observe that distinction. Dretske does believe in purpose and representation of simple organisms and instruments, but he uses them to understand our intentionality without attributing it to them. His theory aspires to establish the reality of *our* mental representation and explain it in materialistic terms.

I do not think Dretske succeeds in either project. The source of the problems lies in the gap between fundamental aspects of mental intentionality and the nature of function and naturalism. The goal of my discussion below is to show that for the purpose of understanding subject-world interactions, our conception of the subject cannot be materialist.

3.4 MATERIALISM

3.4.1 The soprano effect

Dretske's idea of learning is supposed to mark some important distinctions: goal-directed behavior vs. reflexive behavior, internal indicator vs. belief, and tropistic animal vs. a genuine believer. However, it fails. Learning can be found in systems and activities which, by Dretske's own light, are non-intentional. We will use computers as our case study.

Computers are an excellent illustration of the soprano case. They can be conveniently thought of in mental terms. They are programmed to respond to commands such as "open file",

"save", "print" etc. They do not perform any of these tasks because of the meaning of these symbols. The symbols are meaningful only to the users, not to the computer. The machines' behaviors are governed by causal interconnections among hardware elements. The symbols are interface features which facilitate our manipulation of that electronic wiring. Their meanings are useful (though not necessary) for our use, but irrelevant for the computer's behavior. When we ignore the underlying hardware, and focus strictly on the interface features, it will appear as though the computer has thoughts and wants. It thinks I'm trying to save a document. It doesn't want to shut down, etc. But these are merely helpful illusions. The putative thoughts and wants play no causal role in the computer's behavior (82). We read mental representations *into* computers; we do not find them there.

The illusion of computer intentionality is convincing, because the syntactic structure of the software (the symbols) tracks the causal structure of the hardware. The physical maneuvering of symbols mirrors linguistic employment of symbols. "If a symbol's meaning is correlated with the symbol's physical properties --- if the semantics of symbols is faithfully reflected in their syntax … --- then meanings may turn out to be predictively useful without being explanatorily relevant." (81). So mental talk fits computer behavior events because the maneuvering of the symbols and the behavior of the hardware share a syntactic structure. The fit will be as extensive and intricate as the structure.

The degree and intricacy of fit have often seduced us into believing that computers have intentionality. However they only reflect the complexity of the shared structure. Because that structure is shared, or more importantly sharable, by both the semantic and physical employment of symbols, it does not testify for either. Its presence does not indicate intentionality, regardless

of the intentional status of the system. If an intentional system possesses such a structure, it is not intentional in virtue of having that structure.

Dretske's learning is just such a structure. The defining feature is modifiability. A behavior is learnable if its future occurrence can be shaped by its consequences. This is programmable. A computer chess player makes its moves depending on the 'feedback' from its previous move. That is why it appears that the computer is 'thinking'. Modifiability can be replicated at the physical level. Like the hardware of computers, it concerns structural relations among different components and their behaviors (both internal and external to the systems), relations which can be found among objects and states whose causal powers are physical, and do not invoke representational content. Causally relevant content in our intentional states may exhibit such structures. But it is not *in virtue* of such structures that these states are genuinely intentional.

3.4.2 Modifiability and reflexivity

The attraction of explaining intentionality in terms of modifiability, I think, lies in what it denies rather than what it is. Intentional behavior is characteristically not reflexive. So perhaps modifiability can provide a materialist-friendly account for that aspect of intentionality. Why should we be concerned with reflexivity? If an object's automatic reaction to stimuli is sophisticated enough the air of reflexivity vanishes. Computers are a case in point. It does not look like their behavior is reflexive. But as we have seen, that appearance ought not to encourage attribution of intentionality. It is not the absence of reflexivity that matters, but the presence of volition. The contrast should be between voluntary behavior and automatic

behavior. When a system's automatic behavior does not involve modifiability, it looks reflexive. When a behavior is modified, it can be achieved through non-voluntary or voluntary means. Programs modify a computer's behavior without volition. When a person learns to do math, she does so through voluntary exercises of her mind.

Dretske recognizes the significance of volition. Dretske mentions it on a number of occasions as a mark of intentional behavior (109, 122). For example, he uses it as a criterion to rule out homeostatic behaviors (122). However he does not link volition with his account of goal-directed behavior, and, in general, makes no explicit attempt at reduction. Modifiability of learning seems the closest he comes to reducing volition, and it fails.

3.4.3 The materialist's challenge

Here is the challenge for a materialist like Dretske. To explain the essence of intentionality in materialist terms, a materialist must identify some materialist feature which is peculiar to intentional systems, not a feature which is shar*able* by both intentional and non-intentional systems. What can be or cannot be shared is very broad. For example, both non-intentional and intentional animals have eyes. But they use their eyes differently. An intentional animal uses them to decide what to do. Non-intentional animals use their eyes for reflexive behavior. Though intentional and non-intentional animals both share such organs, they do not share their *use* of these organs. We can define an organ by its use, and restrict 'eyes' for intentional beings' use, and call the optical sensory organs of non-intentional beings by another name. In that case, they do not even share the same organs. The difference is terminological. Regardless of which description is preferable, the difference demarcates intentional from non-intentional beings. We

have identified a feature (use of eyes, or 'eyes' in the intentional sense) which is peculiar to intentional animals, and not shared by non-intentional animals[6]. I mean feature in the broad sense.

Any feature we find in existing non-intentional systems clearly cannot be a mark of intentionality, and therefore cannot be the materialist reduction of an intentional feature. Furthermore, if we can construct non-intentional system, even a hypothetical one, involving a certain feature (structure, property, etc.), then that feature also cannot be a reduced version of an intentional property. Now we have a problem. But how can a materialist feature defy being part of a materialist system?

Dretske faces this problem because he draws a strong distinction between intentional and non-intentional beings (believers and non-believers). This distinction calls for non-sharability of intentional properties. If we give up this distinction, we would not need to worry about this non-sharability problem. But then we risk losing a robust, interesting notion of the mental.

Dretske's learning structure fails to identify mental representation. The more interesting question is why this is so. The following examination of this failure will show that some essential features of intentional systems are in conflict with any idea of causal structure as key to understanding intentional representation.

[6] Some philosophers may challenge this claim. But the point here does not bear on how their theories apply to this claim. Here it is merely meant to illustrate how we can think about what is sharable. For that purpose, we can use untutored intuitions.

3.5 INTENTIONAL EXPLANATIONS

3.5.1 Function, causation, and normativity

Learning is supposed to accomplish two things: 1) it shows that representational contents are causally relevant, thereby showing that mental intentionality is not epiphenomenal, and 2) it provides a way for non-derivative normativity to arise out of a set of, otherwise, non-normative objects and relations. In Dretske's view, these two goals are intertwined. Causal relations without normativity cannot establish the reality of mental representation, since representation is normative. Normativity without causation does not provide the right kind of normativity because it does not show that mental representation is not epiphenomenal.

Dretske's idea is that we can find normativity in natural relations such as causation, when causation makes function possible[7]. A non-normative sub-system structure can come to possess normativity in virtue of having a causal role in a system's behavior. Belief is a sub-system state, so function seems a promising way to see how a brain state can acquire normativity. The crucial point here is that normativity depends on causation. No causal role, no function, no normativity.

Dretske's idea of learning uses this insight to show how meaning can arise out of a mere indication relation between a mental state and a fact. To achieve this, he needs to show that the indication relation has a function, by showing that it has a causal role in behavior production. In having the right causal efficacy, it gives meaning reality. In having function, it guarantees meaning's normativity. However, the causal role of the indication relation is questionable.

[7] Function does not always require causation. Stipulated functions, such as those of maps and models, do not. But that kind of function is not useful for naturalism.

3.5.2 Indication and causation

An internal indicator B's indication of fact F is supposed to be causally relevant because it provides a structural cause for the subject's behavior: C→M (where the C is the combination of indicator B and receptor D). The immediate problem is that a structural cause does not provide the kind of causal relation that enables function and normativity. The right kind of causal relation occurs between two events. The rising of room temperature causes the bimetallic strips to bend. The decrease in fuel levels causes the needle in the fuel gauge to move. Dretske's function approach is based on the intuitiveness of how causation provides function in these core cases. Such causation is a relation between events. However, B's indicating F is not an event. It is a general fact about B and F. It does not guarantee that on a particular occasion F's presence will cause C→M. C→M is supposed to happen because M-under-F causes R, which occurs only when F is the case. What the subject needs is an effector mechanism that depends on F's being the case. However, B can indicate F even when F is not the case. Redness in apples indicates ripeness, even if on a fluke occasion an unripe apple is red. The indication fact may mean that there is high probability that this apple at this particular time is ripe. But probability of F, in any degree, is not causally effective. What is required is *the fact that F is the case*. That B indicates F does not guarantee that at the time of the behavior, F is the case. The indication fact may help us predict to what degree F is the case at time t. But as Dretske says, prediction does not have the right kind of relevance (81). So what is needed is for the effector mechanism to be caused by the presence of F. The function of B is not to indicate F, but to be affected by F. When F is

present, B is supposed to be ‘turned on’. The fundamental natural relation is causation, not indication.

3.5.3 Structural explanations vs. triggering explanations

If we replace ‘indication’ with ‘causation’, we must give up Dretske’s way of showing that intentional and scientific explanations are not competitive. He explains this distinction in terms of structural and triggering explanations. In learning, this distinction is applied by treating B as a vehicle carrying content. There is a difference between the vehicle and the internal indicator B on the one hand, and the content that B indicates on the other. The former's causal power is triggering, whereas the latter's is structural. The fact that B indicates F is relevant to M in a way that is distinct from B’s causal role. Unfortunately the relevance is not that of causation. If we give up the indication property, then B is just a member of a causal chain. F causes B to turn on which then causes M, etc. B is not a vehicle carrying content. There is only B, the triggering cause. The fact that F causes B is just another triggering explanation, explaining why B is ‘on’. It does not provide a different *kind* of explanation to why M happened. There is nothing to distinguish intentional explanations from scientific ones.

3.5.4 Causation and explanation

I think Dretske is tempted into thinking that the indication relation can have causal power by thinking about functions in instruments. Because bimetallic strips indicate temperature, they play a causal role in thermostats’ control over the furnace. But what is the nature of this

dependence? The fact that the bimetallic strips are incorporated into the thermostat depends on the creator's *belief* in the indication relation, not on the *fact* that these strips indicate temperature. The fact that thermostats *work* depends on this fact. But it is not right to say that the fact believed *causes* the instrument to work. It *explains* why the instrument works. Likewise, that B indicates F may *explain* the subject's "behavior". However, explanation is not causation. It does not provide a path for derivation of purpose and normativity. Furthermore, the kind of explanation in question here reflects how the theorists make sense of the events, not objective facts about them. In so far as it is dependent on another mind, it cannot explain the nature of mind.

So, Dretske's theory cannot show how B represents F. The representing relationship requires B's indicating F, not F's causing B. If B is merely an effect of F, then its function in the causal chain is merely that of a switch. The causal relationship does not explain how B affects behavior. It only explains when B is 'on' or 'off'. There is no explanation corresponding to belief's explanation of subject's behavior. B's indication of F at most explains behavior in the eye of the theorist. As such it cannot reflect objective facts about B's representational properties, and does not vindicate intentional realism.

3.6 INTENTIONAL EXPLANATIONS AND REPRESENTATIONAL CAPACITIES

Let us return to the origin of our investigations. We begin with two questions

1. What leads to S's action A?

2. In virtue of what is S an intentional system, and S's behavior an intentional action?

Like many philosophers who find intentional explanations a useful starting point, Dretske uses (1) to answer (2): S is an intentional system if the representational content of its internal indicator (and other relevant mental states) are causally efficacious to production of A. This explanatory strategy is based on examining occurrent representational episodes and actions on existing targets.

Question (1) is about a particular event, e.g. Clyde's going to the kitchen at time t. Of course, we are ultimately interested in general claims about how beliefs cause/produce actions. That generalization is supposed to come after we understand how behavioral production works on the occurrent level. For now the focus is on particular behavioral productions.

We tend to focus on actions that are not based on hallucinations. The question begins with a relatively clear description of an action: Clyde's going to the kitchen. Then we ask what causes this behavior? Among the beliefs that govern Clyde's behavior, there may well be beliefs concerning non-existent objects. He may wrongly believe that there's beer or pizza in the kitchen. But Clyde's belief that there is a kitchen must be true. That follows from the description of his action. Then the beliefs and desires will be focused on representations of the kitchen: there is beer in the kitchen; the kitchen is over that way, etc. So such beliefs and desires will not be empty. Empty beliefs are not even candidates for such questions. In everyday discourse, our questions about others' behavior often are not so restrictive. Suppose Clyde believes there is a glass of beer in front of him, and reaches out in that direction, though in fact there is none. We are not likely to ask, "Why are you moving your hand toward that spatial region" (which would focus the representation on a certain spatial region). Rather we would ask "What are you doing?" We would be at a loss in describing his behavior, at least in any meaningful sense. Because there is no clear description of action here, it cannot help the

philosopher begin understanding the relationship between mental states and action.

To summarize, question (1) puts the focus on occurrent exercises of representational capacities, as opposed to representational capacities as dispositions, and the causal role of *non-hallucinatory* mental states, setting aside for later the explanation of hallucinatory ones. From the standpoint of non-empty and occurrent belief-action explanations, the gap between physical causation and intentional behavior production is looking very small. Scientific explanations seem to render intentional ones oblivious. The threat of epiphenomenalism is strongest when our focus is on these cases.

We are clearly aware of some crucial differences between intentional properties and physical ones, intentional non-existence being a prominent one. These features seem mysterious when we take question (1). But they may well testify against (1) as the starting point of our inquiry. The last chapter showed that the real explanatory power lies in intentional properties as *dispositions*, whose exercise on particular occasions may or may not target an existing object. We need to first understand the representational properties of representational capacities, and then understand actual mental events such as seeing, believing in terms of exercises of these capacities. Question (1), however, tries to explain actual mental events first.

I think the explanatory order is wrong. If we begin by exploring representational properties of actual events, we end up with biases towards non-hallucinatory cases. Then intentional non-existence looks mysterious. But in our pre-theoretical thinking about such phenomena, there is nothing mysterious or conceptually puzzling about the possibility of hallucination. One can hallucinate that there is a cat as long as one has the concept of cats (the capacity to represent cats). The aim of the last chapter was to convince the reader that, theoretically, this line of thought points towards the right direction. The heart of the problem of

grounding mental representation is grounding representational *capacities*. Dretske's learning structure provides a helpful start for addressing that problem. A fuller account will be given in the next two chapters.

3.6.1 Function and representational capacities

When a blue jay learns to avoid monarch butterflies, its internal indicator B acquires a new function, the function of indicating monarch butterflies, and becomes a representing state. The job of this state is to "turn on" when monarch butterflies are within the animal's environment. Of course, the state still has that function even when no monarch butterflies are around; the state is just not activated. So what is acquired is a representational capacity, which is exercised when the object represented is present. As we have seen in the last chapter, the capacity to represent is the primary target of our investigation in understanding intentional non-existence. Dretske has provided us with a naturalist account of how an individual acquires a new representational capacity.

Several features of Dretske's account are worth emphasizing. The idea of learning emphasizes *acquisition* of representational capacities. We do not come into the world equipped with an arsenal of representational capacities, just waiting for opportunities to exercise them. That picture makes intentional relations between subject and the world mysterious. If representational relations are to be grounded, they must be grounded in concrete subject-world interactions. In other words, representational capacities must be acquired through subject-world interactions.

For the purpose of our inquiry, acquisition is important not as empirical facts about our cognitive growth, but as facts about what it is for a representational capacity to come into being. How these capacities are acquired should explain essential features of representation, e.g. its normativity, reference, sense, object-independence, etc. The details of this account will be explored in the next chapter. Dretske's account gives us a good beginning.

3.6.2 Accumulation of representational capacities

First, it is worth emphasizing that when we speak of representational capacities, the topic is always capacity to represent a particular affair to state/object. There is no such thing as the general capacity to represent. We do not acquire the capacity to represent the way we acquire the capacity to walk. In Dretske's theory, this is reflected in the fact that B indicates a specific state of affairs, F. When B becomes a representational state, the subject acquires the capacity to represent F, but not necessarily the capacity to represent other states of affairs. To represent G, another indicator B2, and another set of interactions must take place. The general idea of representation or function to indicate is an abstraction from particular capacities to represent or indicate particular things.

Representational capacities are individual capacities that are accumulated throughout the subject's life. The nature and the range of the subject's experience will set a limit on the kind and the range of representational capacities the subject has. This view does rule out robust conceptual holism. However, it does not require that a subject accumulates representational capacities one by one. It is compatible with some form of molecularism. For example, it leaves open the possibility that acquisition of a new representational capacity alters the nature of old

representational capacities. For example, Susan had always thought a table must have four legs. Her capacity to represent a table entails representation of a piece of four-legged furniture. Later she encounters a three-legged object and comes to represent it as a table. Now her representation of tables has expanded to included representation of three-legged objects. There are many much more complicated relations among representational capacities, which we will explore in the next chapter.

3.6.3 Subject-world interactions and normativity

Grounding mental intentionality requires grounding normativity. The normativity of belief owes to its answerability to the world. We want to know how this answerability is grounded in subject-world interactions. According to Dretske's account the representational state B acquires its normative status in virtue of its role in production of goal-oriented "behavior". This is a general fact about normativity of a sub-system part, whether it be bimetallic strips or internal indicators. Their normativity is derived from the normativity of the behavior they help to produce. Whether a behavior is successful or not depends on how well it fulfilled its purpose. Now a believer's behaviors are concrete subject-world interactions. In understanding how normativity of beliefs is derived from that of behavior, we understand how it is grounded in subject-world interactions. Our major task is understanding the nature of this derivation and the source of behavior's normativity.

3.7 NON-DERIVED NORMATIVITY

3.7.1 Learning and purpose

The structure of learning does help us understand the source of mental representation's normativity. According to Dretske, the source of normativity is the goal-directed, or intentional, behavior of the subjects. The goal of the behavior is to satisfy the desire which motivates or produces it. The success or failure of that behavior depends on how well it satisfies the motivating desire. A blue jay eats an acorn to satiate its hunger. If the acorn does the job, then the behavior is successful. If not, then the behavior failed to reach its goal. An organism may discover new desires and hence new goals. A blue jay initially eats a berry to satiate its hunger, and it turned out to be surprisingly tasty. So the behavior is doubly rewarding: satisfaction of hunger, and satisfaction of the desire for tasty food. In the future it will eat berries also to satisfy its hunger.

Representational content is supposed to explain successful behavior. That means it depends on what desire is satisfied. "Why did the blue jay eat the acorn?" ---"It believed that object is an acorn." That does not explain the action, because its representation of the object as an acorn does not show how consuming it would satisfy any desire. What is required is "It believed that object is food." The content must reveal what desire is satisfied. Whether the representation is right or wrong depends on whether the object represented satisfies the desire. If the object did not satiate the blue jay's hunger, then his belief that the object is food is wrong.

This is an overly simplistic and crude introduction of how belief contents may be desire-dependent. However, I do think it is a promising approach. In the next chapter I will argue for this view more extensively. Here I want to point out one important consequence. Representation's normativity, on this view, is rooted in desire-satisfaction, which is in turn grounded in subject-world interactions. This view provides one way to understand how representation's normativity can be grounded in subject-world interactions.

3.7.2 Subject's purpose

The learning structure connects the normativity of belief representation to the normativity of behavior success. This link hinges on the purposefulness of behavior. Here Dretske meets substantial challenges. There are systems which exhibit learning structure and purposeful behavior, but are clearly not believers. The example Dretske offers is homeostatic systems. These systems' behavior seems purposeful. For example, we have a biological thermostat whose aim is to maintain our body temperature within the normal range. It has sensors in our blood stream and in our skin. The sensors tell the hypothalamus of any drop or increase in temperature. It then ignites various behaviors which will retain or dispel heat, such as dilating or constricting blood vessels, shivering, sweating, etc. The resulting temperature is picked up again by the sensors, and cycle continues. Such behaviors are purposeful. Why do Clyde's blood vessels dilate at time t? Because Clyde's hypothalamus 'perceives' high body temperature and wants to bring it down. The explanations mirror our intentional explanations for Clyde's venture into the kitchen. How do we differentiate the two?

Dretske appeals to the nature of their purposes. The purpose of homeostatic behavior is not the purpose of the animal (122). For example, we have a biological thermostat which produces heat generating activities (e.g. shivering) when the brain registers low temperature and heat dispensing activities (e.g. sweating) when it registers high temperature. These activities have purposes, but these purposes aren't that of the subject. When I am too hot, my brain sends out signals which dilate my capillaries. I am not dilating my capillaries. It is an event that happens inside me and happens *to* me, not something I do. Though this activity has a purpose, to lower my body temperature, the purpose is not mine. It is that of my internal thermostat. On the other hand, if I turn on a fan to cool down, this is something I do. The purpose of that behavior is *my* purpose of cooling off my body. My thermostat's behavior is not caused by beliefs or desires. Like computers, it might be convenient to talk as though it is. "My brain wants my body to cool down." "It thinks I am too hot." It is equally superficial. My behavior on the other hand must be explained by my belief and desires.

That there is a distinction between a subject's purpose and a homeostatic system's is clear. But what is the relevant feature of that distinction? Dretske does not say much. He does include a frog's purpose as a case of system purpose. So the idea of system purpose extends beyond believers. Elsewhere Dretske simply assumes that organisms in general have non-derivative natural purposes (64). It's clear that he is thinking of non-intentional purpose, not the kind that Clyde has. It seems that the most obvious candidate here is purpose that is endowed by nature. This notion of purpose, at its pre-theoretical level, is unclear. Theoretically, it has been explained by biologists in terms of natural selection[8]. However, this explanation is undermined

[8] It has also been explained theologically as God's design. But I assume that is not an explanation Dretske would appeal to. In any case, he cannot. According to such explanations, purposes of animal, are like purposes of thermostats, derived from their creators, and hence are not non-derivative.

by Dretske himself. According to him, evolutionary explanations of behavior do not really appeal to a subject's purpose. Evolution tells us why *we find* the behaviors we do, not why the subject does what it does. We find moth's flight behavior because it enhances survival. More specifically, it is the survival of the genes governing this flight behavior. The explanation does not attribute genuine purposes to organisms. In his objection against selection explanation, Dretske has ruled out the evolutionist notion of non-intentional, non-derivative subject's purpose.

There are two more obvious distinctions which are worth mentioning. First, a homeostatic system is a physical part of another system. An animal or a subject is physically independent. But why should physical composition matter? Are Siamese twins one or two systems? If physical integrity is what counts, then each twin is part of a bigger whole, the unified body, and each would be a sub-system part. But the twins are two people who are physically joined, but then do not thereby lose their mental independence. Each can think and to some degree act as independent individuals. The physical conglomerate does not have a mind, and is not a representational system. So, physical independence or dependence is not the relevant distinction between a homeostatic system and its host organism.

A more relevant distinction is the nature of purpose. Homeostatic systems derive their purposes from their host systems, while the purposes of a host system are non-derivative. A biological thermostat has a purpose because maintaining specific body temperature is vital for the organism's survival. In that regard, its purpose is like that of a bimetallic strip, whose function is derived from the thermostat it belongs to.

This way of distinguishing homeostatic systems from intentional systems defeats one explanatory goal of 'learning'. Learning was supposed to show how a system can acquire a kind

of original, *non-derived* representation, or in Dretske teleological vocabulary, non-derived purpose. The defining feature of learning was modification of behavior by feedback. Now we find just such a feature in systems with *derived* purpose. So it fails to distinguish derived from non-derived purposes, and therefore cannot explain the latter.

The structure of learning is not able to identify a crucial element in any teleological theory of mental representation: the source of purpose. Dretske is right to emphasize the difference between the subject's purpose and those of sub-systems. He is also right in looking to the former as the source of purpose that permeates the subject's action and mental states. However the right notion of subject's purpose eludes him because he does not have the right notion of the subject. The subject here is an entity over and above its parts and their relations. Clyde is not just a conglomeration of mental states, the way the joint body of Siamese twins is a conglomerate of the two twins. His purpose cannot be broken down into the relations and purposes of his mental states. The irreducibility of this unity can be seen from the way it is fulfilled. It is fulfilled through Clyde's *voluntary* actions. Volition is a property of the subject, not a sub-system state or part. If the purpose it is used to fulfill can be understood in terms of sub-system parts, then exercises of volition must also be so understood. The latter seems an untenable position. So subject's purpose cannot be understood in terms of sub-system parts. Learning on the other hand is essentially a network of sub-system states, entities and their behaviors which is supposed to endow function onto the constituents, in virtue of this network. That is why it cannot capture the subject's purpose.

3.7.3 Two kinds of holism

The contrast between 'learning' and subject's purpose reflects a distinction between two kinds of holism. Many philosophers accept a certain kind of holism among different mental states such as beliefs, desires and actions. Intentional explanation makes explicit these connections: behavior is made intelligible by the subject's beliefs and desires. When a philosopher takes intentional explanations as a starting point for understanding the nature of beliefs, she takes a holistic approach, and implicitly or explicitly assumes that intentional properties of different kinds of mental states (beliefs, perception, desire, etc.) are mutually dependent and must be understood together. Accepting all this still keeps open two possible directions of inquiry. First, we may understand the holism in terms of interconnections between parts. The internal structures and interrelations among mental states become the primary explanans. This may be the most obvious next step to take, but it is not the only one.

We cannot draw any conclusion about the significance of these interrelations, until we have a good grasp of the elements they are connecting, the mental states. The mere idea of holism is neutral between two ways of thinking about mental states. To speak of the subject's *belief* F is also to speak of the subject's *believing* F. This little grammatical difference, though insignificant in our everyday life, reflects two lines of thoughts with diverging consequences. When our thinking revolves around states such as 'belief', 'desire', we are subject to a tendency to reify these states. We begin to think of them as abstract vehicles with contents. Dretske's idea of internal indicator B is a good example. As a state it is a vehicle, one which may or may not have content, and whose content may or may not have causal relevance. The talk of state and its content is not mere metaphor, for Dretske thinks there is a real distinction between the

causal relevance of the vehicle and its content. The former is shared by both believers and non-believers, where the latter is restricted to believers. Now the way to demystify content is to explain it in terms of relations between vehicles. The subject begins to look more and more like the body that houses these mental states, the way it houses the heart and the lungs.

On the alternative view, belief is something the subject forms and cannot be parsed into vehicles and contents, any more than jumping or falling asleep can be. It is conspicuously bound up with the subject. The subject does not house beliefs and desires, the way it does internal organs. The relations between desires and beliefs are relations between two kinds of things the subject does. They are more analogous to the relationship between standing and jumping, and not the relationship between heart and lung. The heart's pumping and the lungs expanding are doings of the organs. Believing and desiring are the doings of the subject. These mental episodes are brought together by their roles in the *subject's* mental life. On this view, mental holism reflects the unity of the subject, and the unification it brings to mental states. We may crudely contrast the two different types of holism as horizontal connections among states vs. vertical connection between a single subject and her various states.

The holism in Dretske's learning is based on horizontal structural connections. What makes an internal indicator a belief is its role in the causal structure, not its relation to the subject. Without resorting to the relation between the subject and mental states, the causal structure of learning is unable to define an intentional system.

3.7.4 Conclusion

Dretske's theory of mental representation, as a naturalistic reduction, is unsuccessful. He cannot build mental representation out of materialist interconnections of a causal structure. As a structure, it is neutral to both intentional and non-intentional events, and hence cannot define the intentional. Furthermore, the essential features of the intentional, volition and subject's purpose, show that mental states are unified by the unity of the subject, rather than interconnections among mental states. So understanding this kind of mental holism is not a matter of finding the right set of interconnections.

For us, the failure of materialist reduction is not a goal in and of itself. But it does help us answer one of our questions. The question is this: to understand the grounding of mental representation, *must* we conceive the subject and the subject-world interactions in physicalist terms? If there are object-dependent intentional relations, then they may constitute the subject-world interactions which ground intentionality. Then it would not be necessary to begin with a physicalist conception of subject-world interactions, though there is no denying that such interactions are involved. In the last chapter, I had resisted Evans' arguments for object-dependent thoughts, leaving us without intentional grounding.

In this chapter we considered a materialist conception of the subject and interactions between the subject and the world in strictly naturalist terms. It is unable to account for mental representation, because it dissolves the unity of the subject into a set of interconnections. So, important properties which are part of the unity of the subject, the subject's purpose, volition, are lost on such an account. This failure extends beyond Dretske's particular materialist theory. It is a challenge for materialism in general. It also shows that the most basic conception of the

subject must be based on the unity of the subject as expressed through the subject's purpose and volition.

Does this mean that we must return to the intentional conception, which we have seen is also problematic? I do not think so. In the next chapter, I will try to show that there is a middle ground, a conscious, but not (yet) representational system. Such a system is not conceived in mere materialist terms, because it is seen as possessing the unity of a subject. But also, it does not have mental representation. Mental representations are built out of this kind of subject's interactions with the world.

Knowing how to conceive the subject is one step in understanding grounding of intentionality. There is still the question of which subject-world interactions or what structure of such interactions secure that grounding. Dretske's causal structure is actually helpful in addressing this question. Its neutrality to intentional and non-intentional descriptions is a problem for reducing intentionality. But it is not a problem for characterizing intentional interactions. In the next two chapters, I will try to show how such interactions can yield mental representations.

4.0 A PROPOSAL

4.1 THE PROBLEM

Our problem is to understand how demonstrative thoughts' answerability is grounded in subject-world connections. The object a demonstrative thought is answerable to is the object it picks out. Its normative status depends on how things are with *that* object. But not every aspect of that object is relevant to its normative status. Suppose a subject looks at a Victorian era cup and thinks "This cup is white". That the cup is Victorian is not relevant to the truth of the thought. Whether the object picked out is a cup, and whether it is white are the relevant facts. So a thought is answerable to an object *in particular ways*. We expect subject-world connections to explain these two answerability conditions: what a demonstrative thought is answerable to, and how the thought is answerable to it. The relationship between answerability and concrete subject-world connection is made problematic by intentional states' object-independence. Intentional states are answerable to the world even when the objects they are supposed to be about do not exist.

Evans argues that demonstrative thoughts are object-dependent. If they fail to pick out real objects, then they fail to be real thoughts. If he is right, then these object-dependent thoughts can provide the subject-world connections for grounding answerability conditions for object-independent thoughts. However, I think he is mistaken. In Chapter 1, I argued that that

empty thoughts are just as intentional as non-empty ones because they share the same roles in our mental lives. The lesson we should learn, is not that demonstrative thoughts are object-dependent, but rather that we should rethink our conception of what kind of objects thoughts are answerable to, and how to construe that answerability relation.

4.1.1 Object of answerability

An empirical thought is typically taken to be answerable to a *physical* object. For terms, Frege calls the object of answerability 'referent'. For [Evans's] Frege, the referent is the object which makes a sentence true or false (Evans, 8-9). It is the "truth-maker". However, reflection on hallucinatory episodes requires that we widen our conception of the object of answerability. A hallucinatory episode and its veridical counterpart consist of the same thought and have the same answerability properties, though the referent does not exist for the former. Since the absence of the referent does not affect the answerability conditions, its existence is not what is responsible for a thought's answerability. A thought episode, regardless of whether it is hallucinatory or not, does not owe its answerability to any physical relation between object and subject. If it is construed as being answerable to an object, then it would have no answerability. Answerability to physical objects must not be the most basic answerability condition.

The answerability condition that is shared by hallucinatory and non-empty thoughts is that both are answerable to *places*, whether the designated place is occupied or not. It is more fundamental than commitments about objects. So a demonstrative thought owes its answerability to the fact that it makes commitments about places, whether they are occupied by a physical object or not. Commitments about other properties of an object, whether it is a cup or it

is blue, are secondary. An empty thought episode is false in its claim that the designated place is occupied by an object. An illusory thought episode is correct about the place being occupied, but it is incorrect about what properties the object possesses. A veridical thought is correct on both accounts.

Treating places as the fundamental object of answerability shows how object-independence does not jeopardize answerability. But it makes it harder to understand the answerability relations in term of subject-world relations. We want to maintain that how a demonstrative thought is answerable to the world is anchored in some concrete physical connection between the subject and the world. There is no physical connection between a subject and a place. So we cannot understand the subject-world connections underwriting the intentional connection in terms of physical relations. Even when a thought episode is non-hallucinatory, and there is a causal relation between the subject and the object, that causal relation does not ground the thought's answerability. Now thought's answerability to the world seems divorced from any subject-world relation.

4.1.2 Capacity for thoughts

I concluded at the end of Chapter 1 that a thought episode is answerable to the world in virtue of the subject's exercising her *capacity* to have this thought. If she has the capacity to represent an object as a cup, and she has the capacity to pick out a place, then she can form the thought "Here is a cup." On a particular occasion, her ability to exercise that capacity does not require that the thought pick out a physical object. What require grounding, then, are intentional *capacities*.

Grounding is not necessary for every intentional capacity. The issue of grounding arises not because the general ideas of intentionality require subject-world relationship but because, for any particular intentional relationship, its possibility requires some kind of subject-world relationship. The kind of grounding these interactions provide is for *particular* ways of thinking about the world, of being answerable to the world. For example, the capacity to think "Here is a cup" is a particular way of being answerable to the world which differs from the capacity to think "Here is a book." I will call these particular capacities "object-specific capacities".

Object-specific capacities are also capacities to represent objects and properties (e.g. the capacity to represent something as a cup or a book). They correspond to conceptual capacities, or concepts. However, the notion of 'concept' involves much more than object-specific capacities. Concepts are often thought to involve rationality and language. I believe that these features can be set aside to understand how answerability of concepts explains answerability of thoughts. This is controversial. But if this essay is persuasive in what it attempts to prove, then it proves that much of our notion of concept can be understood independent of reason or language. I will use the word 'concept' as denoting the same object-specific capacities to emphasize that the latter is not a different kind of capacity.

4.1.3 Acquisition of object-specific capacities

Object-specific capacities, or conceptual capacities, require grounding in actual subject-world interactions. What object-specific capacities one has depends on what interactions or experiences the subject has had of the world. Our ordinary, non-philosophical, thoughts about concept development support this picture. For example, to have the concept 'cup', the subject

must have had experiences which either directly or indirectly involve cups (e.g. use cups or talk about cups with someone who has the concept). There is a wide array of interactions which directly or indirectly link a subject to cups.

Grounding interactions are not abstract philosophical posits. They are real interactions we find in our lives, namely those which enable concept acquisition. (I will call these 'acquisition interactions'.) In our everyday lives, we predominantly think of these developments as empirical processes, as psychological criteria for concept possession. The philosophical significance of concept acquisition is less clear. These interactions are philosophically significant because they fix the concepts' answerability conditions (that is, how exercises of these concepts are answerable to the world.) As such they are *grounding* interactions. We may seek grounding interactions using our intuitions about acquisition interactions. However we must bear in mind that our interest is their role in the concepts' *answerability* conditions, not their *psychological* conditions.

4.1.4 Conceptual holism

Intuitively, our conceptual capacities are developed by different interactions, and accumulate through our lives as we gain more experience. That requires a certain degree of independence among grounding interactions of different concepts, and therefore a similar degree of independence among those concepts. Some conceptual independence seems undeniable. For example, the capacity to think "Here is a cup" is not the same as the capacity to think "Here is a book." Having the former does not require having the latter, and vice versa. This denies complete holism, but is not incompatible with some kind of molecularism. As Sellars says,

concepts come in batches. One cannot have the concept 'green' without having the concept 'color' or the concept 'object'. Dependence can also occur between the subject's concepts at different stages of her life. In some cases, it is better to think of it as a single concept evolving rather than a new concept replacing an old one. For example, a subject's concept 'grief' acquires a sharper meaning once she actually suffers grief. Prior to that experience her concept is more theoretical (understood in terms of other emotional concepts, e.g. sorrow). Though in a sense her earlier notion is less mature, it is not less genuine. It is just as well grounded. It has its own grounding through other concepts which are grounded in concrete experiences. In general, where there is dependence or independence among concepts, there is corresponding dependence or independence among their grounding interactions.

This view bears stark contrast to a certain kind of holism in Davidson's interpretationism. In Davidson's theory, concepts are attributed according to the best interpretation of the subject's behavior. The legitimacy of the concept attribution depends on the suitability of the theory. Like scientific theories, interpretation theories depend on the amount of data they can explain. Data in this case are subject's interactions with the world. They accumulate in time as the subject gains more experience. With more data, the ideal interpreter revises her theory towards a more correct one. So, the interpreter is in a much better position to provide the 'right' theory at a later stage in the subject's life, than an earlier one.

In an atomist view of concepts, a better theory may mean more concepts. In a holistic view of concepts, a theory attributes all the subject's concepts together in a batch. One cannot add a concept, or revise a concept, without altering the rest. Though revisions for individual concepts may differ in degree, a new theory revises *all* the concepts attributed. Suppose at time t1, the best theory attributes concept 'F' to the subject. Later, at time t2, with new data, the

theorist revises her theory. The new theory would not attribute 'F' but a slightly different concept, 'G'. The claim is not just that at t2, the subject does not have concept 'F' but only concept 'G'. Rather as a revision of the theory at t1, the ideal interpreter is revising what ought to be attributed at t1. It was the best she could do at the time, but in light of further data, it was a *mistake* to attribute 'F' to the subject at t1. This kind of conceptual holism does not just stretch across the subject's concepts at one point in her life. It encompasses the concepts she has throughout her life.

As a theory about what it is to *have* concepts[9], this kind of holistic interpretation conflicts with some of our deeper commitments about the nature of concepts. Part of the nature of a concept is its role in the subject's mental life (including her actions). Presumably it is this connection between concept and action that underwrites attribution of concepts based on behavior. The function of particular concepts in action gives legitimacy to the attribution of these concepts or, if one's not an externalist, to the subject's possession of these concepts. That function is specific to the timing of action. What justify the attribution of possession of the concept 'F' at time t1 are the actions at t1, not the actions at t2. Her actions at time t2 may show that she no longer has concept 'F', but has concept 'G' at t2. That does not undermine the legitimacy of concept 'F' at time t1.

The grounding interactions in my view cohere with the acquisition interactions we find in our lives. The interactions justifying concept attribution in Davidson's theory do not correspond to real interactions. This is a serious flaw, because concepts are defined (at least partially) by

---

[9] It is not clear to what degree Davidson treats his interpretationism as a theory about what it is for a subject to *have* concepts as opposed to a theory about what it is to *discover* what concepts the subject has. As a theory about discovery of concepts, its place in philosophical investigation into the nature of concepts is unclear. In so far as his interpretationism has philosophical significance, concept attribution must amount to concept possession. This is the interpretation of Davidson's interpretationism I am considering here.

their role in real actions, not abstract or philosophically posited actions. If subject-world interactions legitimize concepts, they must be real world interactions. In my case, the grounding interactions aim to justify exercises of concepts in such interactions, and are themselves real.

4.2 GROUNDING

4.2.1 Grounding vs. reduction

Returning to grounding object-specific capacities, our question is: How are object-specific capacities grounded by concrete connections in the grounding interactions? The grounding interactions cannot employ the object-specific capacities they are meant to establish. For example, in the subject's grounding interactions for the concept 'orange', she is looking at an orange. But she is not seeing it *as an orange*, since at that time, she does not yet have that concept. She may see it as a kind of fruit, or a sphere shaped object. These perceptual events presuppose that she possesses other concepts such as 'fruit', 'sphere', 'shape' or 'object'. We will not see clearly the root of concept grounding by looking at cases where an object-specific concept is grounded by interactions using other object-specific concepts. It would be like examining the grounding of a building by looking at how the third floor is grounded by the second floor. Though a building may be multi-storied, the clearest way to discover the principle of its grounding is by looking at the first tier, or even better at a one-floor structure. We will consider grounding interactions which do not use any object-specific concept. Events in such interactions are non-intentional. The goal is to understand how such interactions ground object-specific concepts.

It may seem that we are trying to cook intentionality out of non-intentional events. Our project is aimed at some kind of reduction. I have already laid out my objections against reductionism in Chapter 2. I do not think such projects are promising. How do we avoid reductionism? One aspect of reductionism is not required by the nature of grounding. Reducing intentionality requires reducing an intentional being to a non-intentional being. However, a being who does not possess object-specific capacities is not necessarily non-intentional the way frogs and insects are. Such a being can possess the potential to develop such capacities.

There are some clearer examples of distinct potentials that are required for developing particular capacities. Birds' capacity to fly is learned. A bird is not born with this capacity. The difference between a baby bird and a baby turtle is not that the first can fly but the latter cannot. The difference is that a baby bird possesses the potential to learn to fly. This learning requires certain experiences, usually involving a mother's demonstration. The proper stimulations can induce the chicks to fly. However, no amount of such stimulations can teach a turtle. It simply does not have the requisite potential. It is the presence of the potential to learn to fly that marks an animal as a flying animal.

Similar potential is required for linguistic abilities. Knowing how to speak English is a particular linguistic ability, which must be acquired. To learn a language an animal must possess the general potential to learn languages. That potential marks linguistic beings from non-linguistic beings. Linguistic animals are born with that potential, and non-linguistic animals are not. The distinction between linguistic and non-linguistic animals is not spoiled by the fact that abilities to speak particular languages are acquired, because the latter does not entail that the *potential* to develop those abilities is acquired.

Likewise, we can maintain an irreducible distinction between intentional and non-intentional beings by appealing to a general intentional potential, which enables the development of particular object-specific capacities. That potential is not acquired. More importantly it does not require grounding, because it is not object-specific. The intuition is that for the subject to be able to think "Here is a cup", there must be particular interactions with cups or people who have the concept of cups. We are not trying to ground the capacity to think "Here is ___" independent of any particular object-specific capacities. There is no answerability out there without particular subjects or states being answerable to some *particular* regions of the world in *particular* ways. The working assumption is that *different* answerability relations, or rather different object-specific capacities, are anchored by *different* subject-world interactions. *Specificity* in answerability relations is anchored by *specificity* in subject-world interactions.
The general intentional potential is not object-specific and therefore does not require grounding.

The job of grounding interactions is quite different from the physical events in Dretske's theory. For Dretske, physical events must explain the difference between intentional systems and non-intentional systems by explaining the difference between intentional states and non-intentional states. In my view the difference between intentional and non-intentional systems is explained by the presence or absence of intentional potential, not whether the subject engages in particular interactions. Then what do the grounding interactions accomplish?

Grounding interactions *actualize* the general potential in the development of object-specific capacities. I think it is easier to see the relationship against the background of Aristotle's picture of actuality and potentiality. In *De Anima*, Aristotle describes a three-tier potentiality/actuality structure which fits the nature of intentional capacities (417a21). The three tiers are:

a) First order potential: a potential for having other potentials

b) Second potential/first order actuality: the potential that results for actualizing the first order potential.

c) Second actuality: the excise of second potential.

In our flight example, the capacity to fly is a second potential. It results from actualizing the potential to learn to fly (a first order potential) through the learning process. When a bird is flying, it is exercising its ability to fly, that is, actualizing the second potential. As such it is second actuality.

In the language example, the linguistic potential a human baby is born with is a first order potential. The ability to speak English is a second potential, a result of actualizing the linguistic potential. When a person is speaking English, her speech is second actuality[10].

In the case of intentionality, the general intentional potential is a first order potential. Object-specific capacities are second potential/first order actuality. The object-specific capacity for thinking "Here is a cup" is actualized during a particular demonstrative thought, when a subject is looking at a cup and thinks "Here is a cup". That particular demonstrative thought is second actuality.

These cases can be summarized in the following chart:

---

[10] In the flight and language examples, the animals are born with the first potential, but must learn, or acquire the second potential. I believe this is true for intentional potentialities. However, I do not think it is built into the concepts of first and second potentials that the first is innate and second is acquired. So I offer cases 1 and 2, not as arguments for thinking that general intentional potentials are innate, but as illustrations of how there is a coherent way to think of general intentional potentials as not acquired but object-specific capacities as acquired. Whether general intentional potentials are innate is not crucial to my argument here. What is important is that they do not *need* to be acquired or grounded. If *all* intentional potentialities need to be grounded, then the grounding interactions cannot invoke any intentional potentials, and we would be deriving intentionality from non-intentional potentialities and events. We do not face this problem, since general potentialities do not require grounding.

Table 1: Aristotle's Potentiality and Actuality

Potentiality	Actuality	Case 1	Case 2	Case 3
1^o potentiality		potential to learn to fly	capacity to learn a language.	General intentional potential.
2^o potentiality	1^o actuality (actualizes 1^o potentiality	knowledge of how to fly.	capacity to speak English.	object-specific capacity to have the demonstrative thought "Here is a cup."
	2^o actuality (actualizes 2^o potentiality)	flying at time t.	speaking English at time t.	Thinking 'Here is a cup" at time t.

The role of grounding interactions is similar to the role of fledglings' imitation of their mother's flight behavior. These activities actualize their respective first order potentials (general intentional potential and flight potential). Which parts of the world the subject interacts with and the manner of interaction will determine which object-specific capacity the subject has.

4.2.2. Grounding answerability

Particular object-specific capacities are capacities to be answerable to the subject would in particular ways. The subject can be credited with the first order potential to be answerable to the world prior to experiencing grounding interactions. This potential is not the capacity to be answerable to the world in the manner "Here is ___." The latter is a capacity for a type of answerability relationship, a general one. It is a kind of second potential just as capacities for particular answerability conditions are. The difference between first order potential and second potential is not that the first order is more general, but that the first order is a potential for the second. What I deny is the capacity to be answerable to the world in a general way as a capacity

for a kind of answerability relation between subject and the world as a whole. That relation is not object-specific. All actual answerability relations are object-specific would not be a non-actual, logically postulated answerability relation has no clear place in our theory. What we need is the first order answerability potential, not the second potential of a general answerability capacity.

The subject equipped with first order potential to be answerable to the world partakes in certain subject-world interactions, which in virtue of their non-intentional properties provide the subject with capacities for particular answerability relations with the world. Here are some questions we need to answer. What are the non-intentional properties which are necessary for grounding? That particular physical or causal relations are at play is clear from the fact that these interactions are physical events. What requires further argument is whether sub-intentional conscious properties such as sensation and volition are necessary. There is little doubt that these properties and the first person perspective which comes with them are part of our conception of an intentional being. One may believe that intentionality is inconceivable without them. However, these facts do not show what role they play in *grounding* second-order capacities for answerability. In the following section, I will argue that the nature of answerability requires these properties.

Once we have settled what properties of grounding interactions are at play, there is still the further question of how they lead to second order answerability capacities. Again, we must bear in mind that we are not cooking answerability capacities out of non-intentional events. That the acquired capacities are intentional at all can be attributed to first order intentional potential. It is the specificity of intentional capacities which requires explanation. Specificity of these second order capacities means that the subject can be answerable to different parts of the world

in different ways. The fact that these capacities are acquired piecemeal means the subject does not become answerable to the world all at once, but learns to become answerable to the world bit by bit. So at any stage, there are parts of the world the subject can be answerable to and parts of the world the subject cannot be answerable to. This is the crucial division. We are not interested in the difference between the acquisition of two object-specific capacities the subject *has* (e.g. how the subject acquires the capacity to think "Here is a cup" as opposed to the capacity to think "Here is a book"). We are interested in how it is that the subject has certain object-specific capacities but not others, and how it is that the subject can be (directly) answerable to some parts of the world, but not others. In other words, the grounding interactions need to explain how some part of the world which the subject cannot be answerable to becomes a part of the world which the subject can be answerable to. Now let us turn to the demand the nature of answerability makes on the nature of subject-world interactions.

4.3 ANSWERABILITY

4.3.1 Passive answerability

The way a demonstrative thought episode is answerable to the world strikes many as similar to the way a map or a reading on a gas gauge is. The conditions of answerability are expressed in the content of a representational state. By 'content', I mean what a representational state says. For example, a face on a clock says "It's 5 o'clock." When I am looking at a glass of water, my perceptual experience says "Here is a glass of water." I would say the perceptual content is

"Here is a glass of water." Both inanimate and mental representational contents, what is said, seem to be fixed by intrinsic features of the representational states and the causal events of the situation. Under a certain temperature, the thermometer says "It is 70^{o}F." What it says depends on its intrinsic physical makeup, at that time (e.g. the level of mercury) caused by external circumstances (the surrounding temperature, the level of gas in the tank, etc.). The same seems to be true for perceptual representation. The subject has control over her sense organs with respect to the relevant object, such as turning around, or walking towards it, closing or opening her eyes. But once her sense organs' orientation towards an object is settled, that is once the causal connections between the sense organs and the object are fixed, what the resulting perceptual experience says, the answerability conditions of the experience, are not up to the subject. Both inanimate and perceptual contents' answerability conditions are not up to their respective systems. In that sense, their answerability is passive.

However, there is an important disanalogy. Perceptual states' function depends on the fact that what they say, they say *to* the subject, the system which the states belong to and whose behavior they affect. The subject must register the content of the representation. The perceptual content is what the subject *consciously registers*. That registry[11] is from the subject's first person perspective. (Henceforth I will abbreviate 'first person' as FP.) Perceptual content has a role in the subject's action, belief, in her mental life in general.

Because contents are defined by what the subject registers from her first person perspective, the question "In virtue of what does a perceptual state possess certain contents?" is

[11] Talk of registry is only a rough way of speaking. It has the danger of putting too big a gap between content and the subject. The subject does not register a perceptual content the way she might register the contents of a book. The contents of a book (in typical cases) are not defined by what the book says to one person. They are most likely determined by a larger linguistic community. There can be a gap between what the book says and what the subject registers. However, what a perceptual state says, or means, is defined by what is registered by her in her first person perspective.

more clearly stated as "In virtue of what does a perceptual state possess certain *meaning for* the subject?" How a perceptual state is answerable to the world lies in what it says. So its answerability conditions are defined by what is registered by the subject. With mental states, it is not only that they have answerability conditions, and are first person. They have *first person answerability* conditions. By contrast, artifacts have no first person perspective and cannot register anything. What they say depends on what is registered by other systems (their users) but not by the system they belong to.

Dretske also makes this distinction. He says that the mark of intentional representation is that the representational states say something to the subject, and their content is meaningful *for* the subject (94-5). However for him, the upshot is not that the function analogy is misleading, but that we need to work this subjective element into his functionalist framework. He treats non-FP representation, exemplified by artifacts, as a general form of representation, and FP representation as a particular species. But this is mistaken.

The heart of FP representation is that it is defined by the subject's perspective. Artifacts are not FP, because their representational content is not defined by the system which the representational states belong to. However, that does not mean their representations are perspectiveless. Their representational states depend on the first person perspective of the *users* of the systems they belong to. Hence, their non-FP answerability is derived from the users' FP answerability. FP answerability is not a species of non-FP answerability, but rather its "mother". What we do not have is a legitimate idea of general representation which encompasses both the FP and the non-FP representations.

Dretske's approach of working from non-FP representation exemplified by artifacts to FP representation of intentional systems is a mistake, because he is not proceeding from a general

notion of representation to a specific one, but from 'daughter' representation to 'mother' representation. FP representation is conceptually prior to non-FP representation.

The answerability of perceptual content is passive and FP. Its FP element is not established by first meeting the criteria of non-FP representation. This means that the interactions which ground the capacity for this answerability (the object-specific capacities) directly ground passive FP answerability.

4.3.2 The passive FP answerability problem

There is a problem with establishing passive FP answerability directly. The perceptual answerability is FP in the sense that its conditions are determined by what is registered by the subject. To avoid begging the question, that registration must be sub-intentional. At the sub-intentional level, we experience sensory input and voluntary behavior output. Though we can control the circumstantial condition of our sensory input, we do not have active control of the input itself the way we do with behavior. Since answerability conditions are determined passively by sub-intentional experiences; they are determined by sensations. However, mere sensations cannot deliver answerability conditions in any sense.

Imagine a subject experiencing nothing but sensations. We have such experiences during eye exams when we look into an examining instrument and our visual field contains a series of black letters against a white background. That sensory experience is not perceptual. It does not say anything about how the world is. To deliver any sense of answerability conditions, those experiences must meet two criteria:

(1) It must enable the subject to have the sense that some of her sensory states have normative status. This normative status cannot be true or false, because these are not (yet) perceptual states. Let's just denote the two statuses as +/-.

(2) Furthermore, the condition of what makes a sensory state + or – must be determined by FP experiences. Since we are restricting FP experiences to sensations, the normative status of the sensation in question must be determined by other sensations.

Mere sensory experiences cannot meet any of these criteria. In one's sensory world either a sensation occurs or it does not. There is no issue of good or bad, of success or failure, and certainly nor of truth or falsehood. It is a non-normative event. No variation or complexity can introduce normativity. So, how is passive FP answerability established?

4.3.3 Active answerability

Passive FP answerability of perceptual content is possible because the subject *makes* one set of her FP experiences normative conditions of another set of her FP experiences. The answerability conditions are actively established by the subject. By 'active' I mean 'under the subject's control.' These conditions are established by the subject's voluntary behavior. We find this active answerability when we make commitments about how the world is. In this section, I will try to show the following. First, making commitments is a central part of our mental life. Second, the structure of commitments enables active answerability to bring the world in view. Third, sub-intentional experiences can make commitments possible. Lastly, commitments make passive answerability possible.

4.3.4. Active answerability from commitment

Suppose Mary looks at a cup which appears green to her. But she is not sure the cup is actually green, because she is seeing it under slightly yellow lighting. She wonders whether the cup is green or not. At that moment, she is entertaining the thought "This cup is green" but has not yet endorsed it. When she does endorse it, she adopts the belief "This cup is green". Both the entertaining of the thought and the belief has the same content, the same *passive* answerability conditions. However, there is an important difference. When Mary is merely contemplating whether the cup is green, she has not made a commitment about the world. She cannot be right or wrong. On the other hand, when she endorses the thought "Here is an F", she makes a commitment about how that part of the world is[12]. She is answerable to the world in a way she is not in mere contemplation.

Making a commitment is active. It is an activity that is under the subject's control. This does not mean it is deliberate, or even consciously acknowledged. Most of our beliefs are not results of active decisions. When I look around I 'automatically' believe what I see. - "Here is a table", "That woman is drinking coffee", etc. However, I can revoke these beliefs. "I thought that was a woman. I see now that it is a man." "I'm no longer sure that it was your car that I saw." Our attitudes toward a thought, denial, skepticism, conviction, etc. can all be seen as different attitudes toward a commitment: making a commitment (adopting a belief), refusing to make a commitment (remaining neutral), withdrawing a commitment (revoking a belief),

[12] To make a commitment, in this context, is to endorse some thought content. I will often use these expressions interchangeably with referring to the act. However, they differ in their emphasis. Endorsement emphasizes the relationship between the subject and the content (of perception or otherwise) that is endorsed. Commitment emphasizes the relationship between the subject and the world which she makes a commitment to in a certain way. A subject makes a commitment to the world in endorsing her perceptual content. I want to use both terms to keep in view the triangle between the subject, the thought content, and the world.

denying that one can make a particular commitment (denying that a particular belief is true)[13]. These are different ways of making up one's mind. They are under the subject's voluntary control. Volition is central to active answerability. This is why it is necessary for mental intentionality.

Making a commitment is the same act as adopting a belief. However, more so than the concept of 'belief', the concept of 'commitment' brings out the relations among the thinking subject, the content of her thought, and the world thought about. The structure of commitment shows how the subject's internal world can track how the external world is.

Now we can appreciate another feature shared by hallucinations and their veridical counterparts. They share the same answerability conditions, not only because they say the same thing (share the same content), but also because the subject is making the same commitment. From the subject's point of view, the subject cannot tell the difference (at the moment) between having a hallucinatory episode and a veridical one, not merely because she is confronted with the same phenomenological experiences, but because as far as she is concerned, she is doing the same thing.

4.3.5 The structure of commitment

Making a commitment is a purposeful action, aimed at satisfying some desire (including aversions). The word 'commitment' highlights the fact that this action is the last step the subject can take for achieving her purpose. From there on, the consequences of her actions are beyond

her control. They depend on how the rest of world is. Suppose Jack signs a contract to secure a loan. He is committed, among other things, to pay back that loan. That outcome may not result if circumstances interfere, e.g. the contract turned out to be illegal. These circumstances are outside of Jack's control. Jack's signing of that contract is committal in the sense that it signals the last move Jack can make to affect the outcome.

Suppose Jack's signature alone does not secure the loan. Mary's signature is also required. Jack needs to convince Mary to sign. In that case his signing is not the last step. The "point of no return" happens when he has convinced Mary to sign. In real life, commitments are rarely absolute. Jack has a way to get out. He can break the contract. So it seems that his signature is not a "point of no return." But in so far as we think this is a real option, we do not think his action is that committal. How committal an action is depends on to what degree the action is the subject's last active step in pursuit of a goal.

The aspect of commitment that pertains to us is this interface between what the subject can *actively* do to obtain her purpose, and what the subject must *passively* accept from the world Once the subject's action A is done, whether her purpose is achieved or not depends on how the world is. If A is successful, then the world is such that the subject's purpose P is achieved given her action A. If A is unsuccessful, then the world is such that the subject's purpose P is not achieved through A. The normative status of one's action reflects how the world is objectively. Here is the crucial connection between normativity and objectivity which enables the subject to be answerable to an objective world.

In grounding interactions, this structure occurs at the subject's sub-intentional level. It is composed by purpose-driven voluntary behavior, the sensory input which preceded and led to the behavior, and the sensory input which results from her behavior and constitutes her sense of the

success or failure of her behavior. In her interaction with the world, the subject freely adopts a purpose, and based on her initial sensory input, adopts a behavior as a means to pursue that purpose. In doing so, she 'makes a commitment' about whether her purpose will be satisfied or not. At that point, the success or failure of her goal depends on how the world is. Either the world is such that her purpose is satisfied through her behavior or it is not. This difference is reflected in the feedback sensory input the subject receives as a consequence of her action. With sufficient experience, the subject will associate a particular set of initial sensations (IS) with a particular purposeful behavior (B) and a particular set of positive feedback sensations (FS). FS has normative meaning for the subject which has the form "B is successful". When FS is associated with IS, IS acquires meaning (for the subject) in terms of FS and B. That meaning has the form: "B leads to FS". We might say IS 'says' to the subject "B leads to FS". That content is answerable to the way the world is. Either the world is such that B leads to FS (at the time of B) or it is not. That is an objective fact.

The subjective element shows how answerability conditions are related to the subject. As mentioned earlier, the answerability conditions of a representational state indicate a) which object the state is answerable to, and b) how it is answerable to it. (a) Which object is picked out is here construed as part of B[14]. The object of answerability is picked out by the subject's behavior. The subject's behavior zones in on an object to satisfy its purpose. That is the object of her answerability. (b) The relevant aspect of the object that the representational state is answerable to is determined by what particular purpose is satisfied through a particular behavior. "X is such that it can satisfy purpose P through behavior B." For example, "This is food" can be understood as "Here is an object which can satisfy my hunger by my consuming it." In FP

[14]Later, we will look at how an object is distinguished, and can be construed independently of an action.

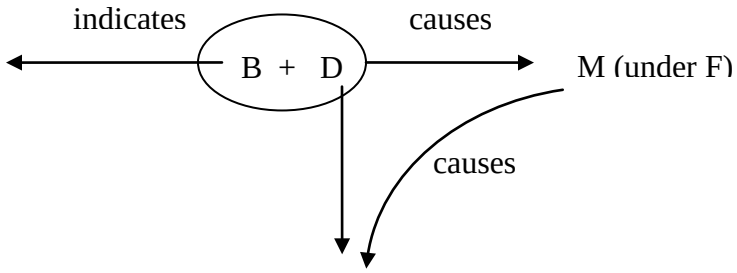
answerability, both (a) and (b) are determined by the subject from her FP perspective. We can now see how. They are results of the subject's active adoption of a purpose and a behavior to satisfy that purpose. The behavior, in so far as it zones in on a particular object, picks out an object as object of answerability. Because the behavior aims at satisfying a purpose, it 'claims' that the object can satisfy that purpose through that behavior.

Determination of answerability conditions is active adoption of purpose and behavior, not passive registration of sensory input. Passive registration defines what sensory input means for the subject. The subjective component (for the subject) can now be understood in terms of purpose satisfaction. A sensory input acquires the function of initial sensory input when it is associated with success of a particular goal-directed behavior. It acquires the meaning "The world is such that the behavior B will satisfy purpose P." A sensory input acquires the function of feedback input on success of B, when it is associated with the latter. Feedback sensory input then passes an internal "judgment" on the normative status of initial sensory input. The active component, the adoption of purpose and behavior injects normativity into the subject's experience.

4.3.6 Dretske's structure

Our structure resembles Dretske's structure of 'learning'. His structure looked like this:

Diagram 3. Dretske's Structure of Learning 2.



Here, "B" stands for internal F-indicator, "D" for R-receptor, "M" for the behavior, and "R" stands for 'reward' which is the consequence of M. At the moment of the behavior, both B and D are present. Together they cause the goal-directed behavior M. M's goal is to satisfy D the internal receptor through consequence R.

Dretske tries to introduce normativity through goal-directed M. When the internal structures are "recruited" into the production of M, they acquire derivative normative status. There is an issue of how they should 'behave,' much like a needle on a face of a clock. Because a clock has a function, to tell time, there is such a thing as where the needle *should* be. Both Dretske and I agree that function is a means of deriving normativity. Furthermore, the feedback loop, that the system can 'register' R, provides the possibility of internal verdict on the systems' indicator. Whether B and D are successful or not can be registered by the subject[15].

The crucial difference between us is that Dretske applies the structure at the physical level, and I apply it at the conscious, sub-intentional level. My view has certain advantages over his reductionist theory. First, the subject has first person perspective at the sub-intentional level. That makes FP answerability possible. As I have tried to argue, intentional FP answerability is

[15] By contrast, in interpretationism the theorist (external to the subject as an agent) determines and registers the normative status of the subject's mental states.

prior to the non-FP answerability of inanimate objects, and one cannot get FP answerability from non-FP answerability. The priority of FP answerability means that answerability begins with FP experiences. Hence the subject-world interactions which establish answerability must be FP experiences.

Second, part of sub-intentional experience is the active adoption of purpose and behavior. This provides a robust sense of purpose which Dretske could not achieve. He had emphasized that the purpose must belong to the subject. Homeostatic systems were ruled out as intentional, because, though they have purposes, these purposes do not belong to these systems. What is for a purpose to belong to a subject in this robust sense? The subject as an *agent* actively adopts the goal. A goal does not belong to the subject in virtue of its structural make-up the way a goal is for a homeostatic system. As I have argued in Chapter 2, internal structure cannot make a system intentional. Intentionality is not a sub-system property. As an agent, the subject functions as a unity which cannot be fully understood in terms of its parts.

4.3.7 Summary

The nature of perceptual content's answerability is passive FP. It is FP, not only in the sense that what it says it *says to* the subject, but that the answerability conditions are *established by* the subject. Ultimately, they are established by the subject's active choice of purpose and action. Therefore what the perceptual content says has a form which relates to the subject's purpose and action, even though it says something about the external world. The general form is "The world is such (at this time) that my (voluntary) behavior B can satisfy my purpose P."

For a sensory input to say this to the subject, the subject must have experienced (that is sensed) that satisfaction of P through behavior B in the presence of that sensory input. So the interaction which establishes the answerability conditions must involve the subject voluntarily behaving to satisfy purpose P under the 'initial' sensations and receive 'feedback' sensations in terms of purpose satisfaction. In time, through association, the initial sensory input says to the subject "The world is such that I can satisfy purpose P through behavior B." Because this is a matter of association, the subject does not have control (at the moment) over what the sensory input says. What the sensory input says, it says *passively*. However, its answerability conditions are established actively in the subject's subject-world interaction.

These grounding interactions enable the subject's otherwise private sensations to say something about the world because the world is "revealed" in the interplay between the subject's active behavior, and passive reception of 'initial' input and 'feedback' input. This is the structure of a commitment. We can say it is 'built' into the content "The world is such that I can undertake behavior B to satisfy purpose P." So when the subject is presented with that content in a perceptual experience, she is in a position to make a particular commitment, to make the commitment that "The world is such that I can undertake behavior B to satisfy purpose P."

To make that commitment is to adopt a belief with that content. It is active. However it is based on the passive content given to the subject in her sensory experience. This is possible because the structure of that commitment is built into that content. The function of perceptual content is to present content for making a commitment. That function is carried out when the subject adopts any attitude toward the corresponding commitment, which includes, merely entertaining it, making it, refusing to make it, etc.

4.4 EXTERNAL WORLD AND EXTERNAL OBJECTS

We have described the general structure of grounding interactions: the subject must behave to satisfy a purpose under certain initial sensory input and receive feedback input in terms of desire satisfaction. Though it is true that this structure makes it possible for the subject to gain 'access' to the external world through subjective experiences, it is not enough. As described thus far, the grounding interactions enable the subject to make commitments about certain *objective facts*, but not necessarily facts about the *external* world.

Imagine the following scenario. You are asked to operate with three machines which are connected to each other. You are asked to look into one machine. It results in an experience like the kind you have in an eye exam, a set of black letters against a white background. Another machine induces a headache. Lastly, your fingers are attached to probes which are registered by a third machine. All three machines are connected, and programmed such that when you have a headache, 'F D G' appears in your visual field, and when you wave your hand, and the headache disappears. You learn that when you have sensations F D G, you can make your headache disappear by waving your hand. In these experiences there is no sense of an external world. Sensations can be used to guide voluntary behavior to satisfy desires without having any sense of external objects. Such interactions would not enable the subject to make commitments about the external world. We have yet to show how such experiences can provide the subject with a sense of an external world.

4.4.1 The necessity for grounding the sense of the external

But does the sense of an external world require the kind of grounding object-specific capacities provide? I had argued earlier that it is *object-specific* capacities, such as "Here is an F", or "There is a G", that require grounding. But commitments about places and objects exhibit an important kind of generality. They seem to be the common denominators of all object-specific commitments. Representations of different places form a unified representation of space, within which the subject can exercise any of her object-specific capacities. Once the subject has a sense of external space, she can make innumerable commitments about specific objects, such as cups or trees, etc. We can attribute to the subject the general potential to represent the external world as part of her general intentional potential, which is actualized by grounding interactions.

The fact that representations of different places are necessarily connected to form representation of space does not refute the idea that the capacity to represent space is built from the capacities to represent, make commitments, to particular places. The fact that there is a single unified space does not mean our capacity to represent it is a single capacity. In fact our capacity to represent space is heterogeneous. Some regions of space can be represented directly, and others only indirectly. The region of space we can directly represent, we can make direct commitments to, fluctuates according to our experiences. Consider the following conceptions of distance.

When I think of where a cup a front of me is located, I think in terms of how I can reach it with my arm, not in terms of how many centimeters it is from me. When I think of how far the

moon is, I think in terms of kilometers from the earth. My conception of an immediate place within my reach is dominated by action. But for other places, my conception is theoretical. That kind of representation involves other representations, those of kilometers, etc. So they are indirect, whereas the former is direct. In investigations of our most basic capacity to represent the external world, we are concerned with direct representations. Indirect representations require advanced conceptual capacities which presuppose the basic representational capacities we are investigating.

Space of which we can exercise direct representation is what Evans calls "ego-centric space". I had suggested that we should think of that 'space' in terms of a *conception* of space. Evans draws a strong connection between ego-centric conception of space and behavior[16]. It is about what kind of *meaning* places have for me, not about what kind of places they are. When a place has ego-centric meaning for me, I see it as one in which I can act. This is the main difference between my conception of the distance of my cup and that of the moon. An object right in front of me is within my reach, I can (and usually do) think of it ego-centrically. The moon is not within my space of action, so I cannot have an ego-centric conception of its place.

Our conception of space, what kind of meaning a particular place has for us, or if it has any meaning at all, is acquired through our actions, and are grounded in them. Actions are meaningful to the subject only in so far as they generate feedback that is registered by the subject. Only actions on objects, not actions in space, can generate such feedback. So, our conception of external space is acquired with our conception of external objects. The grounding interactions must secure both. The question is: What aspects of the grounding interactions provide the sense of external space and objects?

[16] "[W]e must say that having spatially significant perceptual information consists at least partly in being disposed to do various things" (155).

4.4.2 Distinction of self vs. the other

The idea of self vs. other which I am trying to obtain here is relative to objects and places. There is no self vs. the other that is established "once and for all." The meaning of externality for new spatial regions and objects occupying them must be gained piecemeal. The grounding interactions must provide the distinction of self vs. the other with respect to particular objects and space of the interactions. This is possible because part of the feedback the subject receives from her behavior is resistance to her volition. For example, when she reaches the cup, there is resistance to her moving further. She cannot occupy that region of space without moving the cup. In our daily lives, this kind of feedback can be used to discover objects in one's environment. Imagine someone fumbling around in the dark to find a wall of the room she is in. She has no perceptual input for her guidance. Her action is unguided, though voluntary. When she encounters a wall, she feels the resistance to her action, as new perceptual input. She gains a sense that she has encountered an external object. When there is no such resistance, the subject gets no sense of external objects from her sensory field. The relation between sense of resistance and externality is well illustrated by 3-D movies. When watching these movies, our visual sensations tell us that there are real objects in front of us. But when we reach out to touch the presumed object, there is nothing to act against, nothing which gives check on our volition. There is no resistance where we would expect there to be. Through such scenarios we learn that these 'objects' are not within our space of action. So our actions, not mere sensory inputs, tell us that there are in fact no objects.

Mere sensations whatever the quality (inner or outer) cannot provide a distinction between self and other. But when sensations are used to register the difference between the limit and the freedom of one's volition, they enable the subject to sense the distinction between self and other. The externality of objects is sensed as externality to one's sphere of volition. The passive feedback on active behavior gives rise first to the distinction between self as agent, and the object as patient.

Just like with the types of feedback sensations discussed previously, the meaning of this feedback for the subject can be attached to an initial set of sensations, when it is associated with the latter. When these 'resistance' sensations map onto initial sensations, the subject learns how to 'interpret' her 'initial' sensations as providing information such as "here is an external object," "that is empty space." For example, when I look at a cup, I initially receive initial sensations from the cup. These sensations themselves cannot tell me that they are from an external source. They may just as well be hallucinations. When I reach out to the cup and feel the resistance of the object, the point of that sensation matches with parts of my non-feedback initial sensations, the edges of a white shape, for example. There is no check on my action in the visual space next to that white shape. So I know where the object ends. Suppose there is a black circle on the cup. I can tell it is part of the cup because I can correlate that sensation with the resistance I feel when I touch that area of the cup. But I do not register it as an additional physical object, because I feel no difference in my resistance sensations, e.g. those I may feel if the cup has a hole or the area is protruding. The interface between freedom and limit of my action enables me to register the existence, the shape, the size of an external object. Consequently I can delineate objects through my perceptual experiences.

Part of this achievement is my ability to register places in my ego-centric space as occupied or empty. The sense of externality from an object gives me the sense that my space is not a sphere of my inner world (for example, a world of floaters), but an external space, one in which I can relate to other objects through my actions. In time these experiences enable the subject's sensations to acquire new meaning for her. She interprets her sensations as awareness of an external object.

The coordination between passive sensory input and active behavioral output in grounding interactions enables the subject to see an object *as a part of her external world.* It is *hers* in the sense that she can relate to it through action. It is part of her ego-centric space. That ego-centric space is *external*, in the sense that occupants of that space put a limit on her volition. So, though we can take the existence of space for granted, we cannot take for granted the existence and expanse of one's ego-centric space. In other words, we cannot take for granted what regions of space *mean* objective space *to* the subject. For the subject to directly act toward a place, the place must not only exist but must be *hers*, that is, be part of her ego-centric space. She can only act in places which she has learned to be her ego-centric place. Part of the converse is also true. For her to act towards a particular place, she must 'see' that place as a place of possible action. The notion of 'seeing' is not a psychological one; it is true by definition: ego-centric space is the space of possible action. Her ego-centric world is a world of possible (direct) targets for her commitments. In other words, it is the world of *possible* direct references. (Indirect references would expand from here on through other representations, e.g. concepts.) Grounding interactions expand the parts of the objective world as possibilities of direct reference for her.

4.4.3. Movements

These experiences also give her (voluntary) behavior the significance of acting in an external space. Now when she reaches out, she can see her movement as movement in public space, as reaching out to external places. There are voluntary movements whose feedback does not enable us to locate in external space, e.g. swallowing or flexing one's muscles. We know these are movements happening in an object occupying a public place because of our knowledge of ourselves as such objects. But from the agent's point of view, the check on the subject's volition does not map onto any feedback sensation. Such behavior cannot be located in the subject's ego-centric space.

The grounding subject-world interactions enable the subject's sensations and voluntary movements to acquire their meaning of externality in coordination. However, the interpretation of sensation as awareness of external objects and that of one's voluntary movements as movements in space are distinct. In practice, one can see oneself as moving in external space without having sensations which one interprets as sensations of external objects. Likewise, one can take oneself to be seeing external objects without experiencing herself as moving in external space.

The externality of objects, places, and behavior are one. Only actual objects can give the subject resistance feedback (in fact any kind of feedback), therefore only objects, not places, can enable the subject to have a sense of the externality shared by both objects and space. The acquisition, the *grounding*, of the subject's capacities to make commitments about places and object in the external world is object-dependent in this sense. However, the *exercises* of those capacities through sensation or movement are object-independent.

4.5 CONCLUSION

4.5.1 Summary

Let us review the nature of making a commitment in demonstrative thought and how that capacity is grounded. A subject has the object-specific capacity to make the commitment "Here is an F" when a certain set of sensations means or says "Here is an F" to the subject. This is the perceptual content 'presented' to the subject for endorsement or commitment during a demonstrative thought episode. Its function is to "suggest" to the subject what commitment to make. That is, perceptual content is essentially an object of commitment. We can think of the content as essentially 'primed' for making commitments.

It is so primed because it is grounded by the interplay of passive input and active output that constitute the structure of a commitment. The grounding interactions consist of initial sensory input, active behavior, and sensory feedback. For the subject, her association of these three elements enables her to use those initial sensory experiences to satisfy her purpose. In her pursuit of her goal, her action is an act of a commitment. Once the action is performed, whether the subject's goal is met or not depends on how the world is. In choosing to undertake behavior B to pursue purpose P on a particular occasion she makes the commitment "The world (now) is such that my purpose P can be satisfied by behavior B." So the association between the three elements of the grounding interactions enables the subject to make that commitment about the world based on her initial sensory experiences. In other words it grounds her object-specific

capacity to make the commitment "The world is such that my purpose P can be satisfied by behavior B."

As ground for a commitment, the experience says to the subject "The world is such that my purpose P can be satisfied by behavior B." The grounding interactions must confer that meaning onto a set of sensory experiences for the subject. This involves mapping feedback sensations onto those sensory experiences which function as 'initial' sensations. In a particular grounding interaction, there are many aspects of the feedback. Feedback as resistance to her volition enables the sensory experiences to signal to the subject the presence of an external object. The differentiation between resistance and absence of such resistance enables the subject to delineate objects with those sensory experiences. Where that resistance feedback comes from allows these experiences to locate the object for the subject in her ego-centric space. The purpose-satisfaction feedback the subject experiences as part of that feedback gives the sensory experience the significance of signaling to the subject a situation where purpose (typically a desire) can be satisfied by a particular behavior. When the desire-satisfaction is mapped onto the volition resistance sensation, the experiences are able to say "Here is an external object at this location such that my behavior B can satisfy my purpose P." The meaning of these sensory experiences (i.e. the content of perceptual experiences) is registered by the subject terms of action and purpose. There is a psychological connection between having such experiences and thinking certain actions are legitimate. However, this connection is not merely a contingent psychological fact. It is part of what it is to have perceptual experiences.

Initial sensations can carve out the world into meaningful objects because the grounding interactions involve behavior whose desire-feedback is correlated with the initial sensations. Which sensations, which behavior and which desire or resistance feedback, and how they map

onto each other determine the specificity of an object-specific capacity. These sub-intentional events are underwritten by physical interactions between subject and world. So the particular physical interactions ground the particular commitment capacities the subject has and explain why parts of the world have meaning for the subject and others do not, and what meaning they are.

4.5.2 Object-independence

Whereas a commitment is a relationship between the subject as an agent, her perceptual input (or meaningful sensory input) and the world, the meaning of a sensory experience is just a relationship between that sensory experience and the subject (as an agent). An experience says such and such to the subject. Once the meaning is acquired, how that experience is related to the world does not affect the meaning. If an experience says "Here is an orange" it will say so whenever it is presented to the subject, regardless of whether it is a result of the subject's looking at an orange, a hologram, computer manipulation of the subject's brain, or any other cause. The meaning also does not depend on the subject's exercise of volition at the moment of her sensory experience. Whether or not she endorses the content, chooses to act on it, things will *look* the same way to her. A hologram still *looks* like an object floating in space even if the subject knows there is no object there. That is, her perceptual experience still says "There is an orange there," even though she does not endorse it. Though the meaning of sensory experience is conferred through active behavior in grounding interactions, it is passively given to the subject in perceptual experiences. It can be altered by future interactions, which ground (and impart) that change. For example, after some time, a subject might learn to see a mirage as a mere visual

aberration rather than a pool of water. Not only does she know that there is no water, but it does not even *look* to her as though there is. Her experiences with mirages ground the *loss* of the intentional content. The sensory experience no longer means that there is an external object.

The way to understand the passive elements of sensation and perceptual content is through commitment. Sensations provide the object of perceptual meaning, what the meaning is attached to. The resulting perceptual content provides the object of commitment, what the subject endorses. The role of sensation is to provide the FP ground for the subject's commitment, not a causal connection between the subject and an external object. That is why how the sensation is caused does not matter. During the thought episode, the sensory content's job is to enable the subject have meaningful sensations. That is all it is to have perceptual content. That only requires that the subject possesses an object-specific capacity which features that sensation, which means the subject attaches meaning to the sensation. The meaning of the sensation is grounded in the grounding interactions in the subject's past, not in subject-world interactions of the demonstrative thought episode. That is why the absence of subject-world interaction does not affect the subject's ability to possess the passively answerable perceptual content and to make a commitment with that content.

At the heart of that commitment is the content in the form of "Here is an object which can satisfy D through action A." This capacity is developed from grounding interactions which, in the case of intentional beings, actualize the general intentional potential, enabling sub-intentional sensations and movements to acquire meaning for the subject. The sensations of the perceptual content, the basis of the commitment, must say to the subject "Here is an object which can satisfy D through action A." They must signify an external world, which is possible because the

grounding interactions involve coordination of sensations and voluntary behavior which enables initial sensations to map onto resistance sensations.

4.5.3 Reference and sense

Making a commitment involves picking out an object one's commitment is about, answerable to, and what commitment one makes about that object. It is active because both aspects of the commitments are determined by the subject's choice. Having a sense of external objects and space enables the subject to pick out referents (object of commitment) within that space. The subject relates to objects and places (directly) through action. It follows that the subject picks out an object through action. For example, the fact that she is targeting a particular region is manifested by her reaching out to that spot and trying to pick up an object. In grounding interactions, action requires feedback from objects to provide meaning for the subject's sensations and *establish* the capacity to use that meaning in action. The actual *exercise* of that capacity in action does not require that there be an object which functions as an object of action. Action necessarily takes place in space, but it does not necessitate the existence of an object. The intended target may be an object, but the target the action is necessarily answerable to is first and foremost a place. As we said in Chapter 1, that the most basic commitment is commitment to places. Whether a place is occupied by an object, or even occupied by an object for her, does not affect her ability to (directly) target the place. Commitments about objects are not "Russellian". However, if she makes a commitment about a place, then that place necessarily exists for her. That is she can legitimately think "Here" with respect to that place. All commitments imply a legitimate use of "Here". Therefore here-thoughts are Russellian.

Through voluntary behavior in grounding interactions, the subject gains the capacity to use her sensory input as awareness of external objects and her behavior as action in external space, and hence the capacity to make commitments about external objects. Her ability to target external objects is not sufficient for making a commitment. Making a commitment is rooted in purposeful action. To make a commitment towards an object, the subject must see the object as relating to her present purpose. We have already discussed how the grounding interactions through desire feedback provide the subject a way of seeing the world. "The world is such that I can satisfy my purpose P with behavior B." Now that we see how grounding interactions also enable the subject to delineate the world into objects, we can fine tune how the subject sees the world. She sees the world as populated by objects which have meaning for the subject in the form of "This object is such that I can satisfy my purpose P with behavior B."

The subject's capacity to pick out an object and her capacity to see it as such and such are inseparable in making a commitment. In making a commitment, the subject acts towards an object to satisfy a purpose. In her action, she picks out an object among a world of objects. How fine-grained her action is, like its other aspects, is determined by her purpose. If her purpose is to climb a tree, then she picks out a tree as object of her commitments. If her purpose is to examine the health of a tree by looking at its bark, then she picks out particular pieces of bark as objects of her commitments. How a subject picks out an object is an integral part of how she sees an object. In grounding interactions, both are established in the same behavior (type), because in the feedback of that behavior, she feels the resistance to her volition, and the satisfaction of her purpose. The first enables her to see that her purpose is satisfied by an external object, and the second enables her to fine tune her sense of the external object to relate it to herself.

So though there is a conceptual distinction, there is no separation in establishing the subject's capacity to refer to particular objects and to represent them in particular ways. In other words, there is no separation in establishing reference and sense. There is a legitimate distinction in the sense that in practice, the same object can be represented in different *ways*. But what counts as an object is itself a *way* of representing the world. The distinction between the target and how it is seen by the subject is legitimate only in the limited context where the target has been stipulated as fixed. Then we can say such things as Jack sees this ball as a toy, but Joan sees it as a weapon. They represent that object in different ways, but share one aspect of their representation of the world: it contains this object.

4.5.4 Narrow vs. wide content

The distinction between representation-of and representation-for lies at the heart of another, better known, distinction: narrow vs. wide content. The ideas of narrow and wide contents developed from consideration of mental contents of *rational* animals. But the work these concepts do does not turn on the conceptual aspect of our mental content.

In "The Meaning of 'Meaning'", Hilary Putnam argued that our mental content is partially determined by our external environment. In his Twin-Earth thought experiment, two worlds, Earth and Twin-Earth are identical, except that instead of H_2O, Twin-Earth has the substance XYZ. XYZ has the same physical and chemical properties as our water. It looks clear, tastes bland, boils at 100^oC, it is even called 'water' in Twin-Earth's English. Oscar is on Earth. He has a doppelganger on Twin-Earth. Both interact with their version of water in their daily business. They can't tell the difference between the two substances. In fact, if we isolate

their mental history, we couldn't tell which one belongs to Oscar and which to his doppelganger. They treat their 'water' in the exact same way, since both substances have the same apparent physical properties, and figure into their lives in the same way. Whey they say to themselves "I want to have some water", do their thoughts express the same content? Putnam's intuition is 'no'. They are in fact thinking about different objects. When they say, "I want water" they want different things. For Oscar it is H_2O, for his doppelganger it's XYZ. A number of people were convinced that content is partially determined by external environment, not just by how we see the environment. This conception of content is called 'wide content.'

The 'wide content' view is right in that the mental content of "water" in his example is a representation *of* an external object in the subject's environment. Oscar and his double ganger's representations are of different objects. Though Oscar and his double ganger's thoughts are both described as "I want water" what they want are different things. One wants H_2O; the other wants XYZ. Hence their desires provide bases for different actions on different things. Therefore their representational contents are not the same. "Wide content" is right in capturing the representation-*of* component of mental content.

However, 'wide content' ignores an equally important aspect of mental content: how a subject represents objects. We are tempted to say Oscar and his double ganger have the same mental content, because they represent their referents in the same way. They both represent them *as* water. XYZ means the same thing to the double ganger as H_2O means to Oscar. The inability to explain representation-as does not leave the theory incomplete, but null. If S represents X, it necessarily represents it as something. A theory which only focuses on representation-of is not a half theory any more than one face of a coin is half a coin. So while

Putnam's argument highlights the representation-of elements of our mental content, how it figures into a coherent theory of mental content is yet to be seen.

Conceptual role semanticists swing to the other end. Their views are best seen as explaining the representation-as element. Ned Block suggests that mental content should be defined by its causal role in its interaction with other mental contents and as mediations between perception and action (p. 93). The emphasis is on content's role in the subject's internal world. The causal relation between mental content and world is excluded. “Conceptual role abstracts away from all causal relations except the ones that mediate inferences, inductive or deductive, decision making, and the like” (p. 94). These interactions seem to show why both Oscar and his double ganger see their referents as water. Their representations have the same role in their mental lives.

However, an exclusive focus on interconnections among mental states cannot explain how objects are meaningful for the subject. Conceptual role illuminates the nature of a representational content by placing it within a sea of other representational contents. It does not show how there can be representation in the first place. It cannot not show how any of these representations are grounded. I have argued in chapter 2 that no structure however complex can deliver meaning. Conceptual role semantics is one theory which draws on internal structure to ground meaning[17].

An adequate theory of mental content must show not only how it embodies representation-of and representation-as elements, but how S represents X as Y. In the simplest cases I have been considering, what determines that S's representation is of X is the fact that his (intentional) actions home in on X. S always acts on an object as an object which would satisfy

[17] Other theories include functional theories (Block 1986, Harman 1973)

his purpose. Representation of X and representing it as Y are inextricably bound together by the homing in of action and the purposeful nature of action. The basis of action, of course, involves more than desires. It involves the subject's cognitive perception of the world, e.g. there is a nut here. In more realistic, and more complex, cases, both the cognitive and the non-cognitive worlds form a vast web. Despite the complexity, the essential structure is still the same.

4.5.5 Direct and Indirect representations

Thus far we have only considered direct representations, in which neither reference nor sense is grounded through other representations. Direct reference is grounded in the subject's homing in on a particular object. Direct sense is grounded in the subject's perception of that object in terms of feedback from her actions (especially in terms of purpose fulfillment.) In turn the actions must be perceived as direct by the subject. I think Evans gives us the best way to think of direct action, action in ego-centric space. When we conceive of an action as 'up', we are not thinking about it in terms of the representation 'up', as we are when we think of actions as "going west 1000 kilometers from here." Rather the notion 'up' is possible because we have a practical grasp of such behavior. 'Up' and 'here' are based on our practical, non-descriptive, conception of ourselves. The ego-centric framework defines the spatial territory of direct representation.

Just about everything we do is either an action in ego-centric space or conceived in terms of such action. For example, sleeping, watching TV, jogging, brushing one's teeth, reading a book, finding a building, etc. These activities involve complex representation, but are direct actions. They make possible complex indirect representations.

Indirect representations are grounded in subject-world interactions through other representations. Ultimately, all are anchored in the grounding of direct representations. The structure is much like a multi-story building, in which the first level is directly grounded in the foundation, and the upper levels are grounded in the foundations by being grounded to the floor below. The possibility of mediated grounding enables a rich and complex stock of representations. There are different kinds of mediated grounding. Here I will focus on the grounding of descriptions.

For example, my conception of 'a kilometer' is based[18] on my conception of 'a meter', which is in turn based on my conception of space as 'this far from me', an ego-centric conception. While the idea of 'this far from me' requires action which involves 'this far from me', my conception of 'a kilometer' does not require, and is not grounded in, actions involving a kilometer, e.g. traveling a kilometer. That conception is grounded in my conception of 'a thousand' and 'a meter'. (The former representation is more complex, and we will ignore that for now.) In that sense it is indirect. One the other hand, even if one travels a kilometer numerous times, the representation may still be indirect if it is conceived in terms of a series of 'this far' representations. For one's conception of a kilometer to be direct, one's ego-centric space must extend a kilometer.

The distinction between direct and indirect, in terms of grounding, is not a difference between different concepts, but between different *conceptions*. Though there is something to be said for that difference (which will be said later), we should observe that in ordinary language, the distinction is blurred. For example, one biology professor once said to me "A cell biologist's

[18] I say 'based' not 'derived' to emphasis that it is the legitimacy of the concept, not the psychological possibility of the concept, that is at issue.

concept of ‘life’ is different from a geneticist concept of ‘life’.” This is a difference in conception of life. These conceptions are reflected in the nature of a judgment involving these concepts. For the cell biologist, the claim “A living cell must exhibit movement” is (almost) true by definition. But for a geneticist, it is a contingent claim. So, we make judgments using our conceptions. That is a vital role which we often reserve for concepts. It is helpful to realize that often conceptions play this role.

Our talk of the distinction between direct, non-descriptive, practical representations and descriptive representations can be misleading. The distinction is really the distinction between two modes of *grounding*. The foundation of representations of the same objects can change from one mode to another. Suppose Joe has never seen snow. Mary describes snow to Joe. “Snow falls from the sky. It forms a white blanket on the ground, and can melt under the sun.” Joe now has a descriptive, indirect representation of snow. When Joe moves to a cold climate and begins to live with snow during the winters, his conception of snow is now governed by his activities and direct observations. It becomes more practical, and less descriptive. For example, after working with a lot of soiled slush, his idea of snow is no longer governed by the idea that it is white on the ground. Being white may be characteristic of snow, but not an essential property.

Ideas can also change from practical to more theoretical. For example, prior to any education in chemistry, Fred thought of water only in terms of everyday activities, it’s clear, drinkable, non-poisonous, etc. But after he has studied the chemical properties for a long time, the fact that water is H_2O encroaches upon his conception of the substance. So though once upon a time the claim “Water is H2O” was a scientific discovery for Fred, it is now more akin to an analytic truth. His conception of water is less practical and more theoretical.

4.5.6 Concepts in general

We have used demonstrative thought episodes, in particular hallucinations, to address the questions: How can object-independent intentional states be answerable to the world? How can such answerability be grounded in subject-world interactions? How does the world become meaningful for the subject? According to my theory, it becomes meaningful through its role in the subject's world of purpose and that the answerability of our conceptual capacities is grounded in our purposes and our attempts to use the world to achieve them. How well can this purpose-based theory apply to a greater range of concepts?

The structure is not confined to the sub-intentional level. We focused on the sub-intentional to understand the direct grounding relationship between conceptual capacities and the world, unmediated by other concepts. I had used the building analogy to illustrate what we are trying to accomplish. To understand how a building is supported by its foundation, it is best to focus on a one-story building. The principles we learn from that investigation are applicable to multi-story buildings, once we understand how one floor can be supported by another. Likewise, the general structure put forth is applicable to higher level concepts. The general point is that voluntary behavior and sensory awareness utilized to achieve purpose bring the relevant part of the world into the subject's mental world. This is true even if the behavior is action, and the sensory awareness is perception, and the purpose is an intentional desire. Some of the examples given are of this kind. Suppose the subject is looking at a mirage, but she thinks she is looking at a pool of water. She is thirsty and with no other sources of water in sight, she walks towards the mirage to drink from it. Her desire, perception, and action are all intentional. They exercise her concept of water, thirst, etc. Now as she approaches her object of desire, her experiences will

give her perceptual awareness a different meaning of what she is seeing. She will realize that she does not get closer to the object, but rather the object becomes further as she walks. She learns that she is not looking at a physical object occupying a fixed place. Now the same perceptual experience will mean to her a visual aberration rather than a physical object. So the same structure can be instantiated at higher levels.

In the next chapter, I will explore in greater depth how the grounding interaction structures can be expanded and held accountable for complex representations. I will also show the resulting picture of mental content and how it relates to some contemporary debates.

5.0 CONCLUSION

I have argued that representational capacities are grounded in intentional subject-world interactions: perception, action, and perceptible feedback (in terms of sensation and desire). The particular representational capacities are individuated by the nature of the sensation, action, and the resulting feedback. An animal can represent something as food, if it knows (practically and implicitly) that its hunger can be satiated by consuming objects. To have the capacity to represent objects as food is to have this practical knowledge. On any particular occasion, the subject actually sees something as food, if she believes that her hunger can be satiated by consuming that object. This subject-world interaction structure lays out what it is to have these capacities. This structure is a general framework, which I will call 'the intentionality framework'.

5.1 PURPOSE

We have been considering the most primitive application of this framework. But it can be enriched in many ways. First, purposes go beyond simple desires. They include any want, need, interest, concern, or goals. We have considered non-intentional primitive desires provided by nature. Once the subject develops representational content, she will have intentional desires. Such desires can also fit into the intentional framework, and produce more representational content. These first-level intentional desires will spawn derivative desires, such as the desire for

money to get food. Extension of such derivation can lead to complex and advanced purposes. Some representations rest on multiple desires, because they are derived from other multiple representations. For example, a cup is a drinking vessel. The subject must have the idea of what it is to drink, and the idea of a container. That means the subject must have desires fulfilled by drinking (thirst) and other desires whose fulfillment requires objects containing other objects, for example the desire to store grain. These purposes form a rich network which shapes our representations.

5.1.1 Purpose and the subject's perspective

A subject's purpose is the primary first-person feature which defines the subject's perspective. We have seen the significance of first person perspective. What an object means to the subject is governed by the subject's perspective on the object. The subject's actions are based on how the subject views the world. This needs to be understood in terms of the subject's purpose.

There is another kind of first person's perspective which features more prominently in the literature: the subject's raw phenomenological experience, etc.[19]. By phenomenal experience I mean the passive experience of sensations. This kind of first person is non-intentional. Sellars has argued that because it is non-intentional it cannot function as foundation of knowledge. The distinction Sellars was concerned with is that between nature and reason. Mine is one between nature and intentionality. But his point is applicable. In fact, I think the gulf between nature and reason is due first to the gulf between nature and intentionality. (See section 2, for a discussion

[19] See Loar (1999), Bermudez (2000)

of the relationship between intentionality and rationality.) Raw sense data do not define what an object *means* to the subject; they merely make the object accessible to her.

5.1.2 Non-intentional phenomenal experiences

We need to distinguish between intentional and non-intentional first-person. The first only actualizes the potential for consciousness. The second actualizes intentional potential. Philosophers have used the concepts "consciousness" and "perception", "awareness" in both intentional and non-intentional contexts. This ambiguity leads to debates about whether we directly perceive sense-data (McDowell 1994, Dancy 1995), and indirectly perceive external objects, or do we directly perceive external objects (Russell 1912, Jackson 1977, Robinson 1994, Ayer 1958, O'Shaughnessy 2003). In this debate, the primary difference between awareness of sense-data and awareness of external objects is a matter of directness. But these are two different kinds of awareness which are not on a par at all. The first is non-intentional awareness, and the second is intentional awareness. We may use the word 'perception' broadly enough to cover both kinds. Then we should distinguish between non-intentional perceiving and intentional perceiving. The latter is not derived from the former, because intentionality cannot be derived from non-intentionality. As for directness, both are direct. Phenomenal awareness is direct non-intentional awareness or perception if you like, and awareness of objects is direct intentional awareness or perception. An indirect intentional perception would be e.g. seeing a brick as a door stopper.

On the other hand, Robert Brandom (2002) has argued against the significance of non-intentional phenomenological experience. According to him, phenomenology primarily

functions as information channels in our intentional experience. Information channels can extend far beyond phenomenology. Telescopes, microscopes, MRI, all provide information about the world. They make it possible for us to 'see' micro-level, or distant objects which are not available to the naked eye. Though such mediated perception is causally dependent on phenomenological seeing, it can play the same role in our cognitive lives no less. As far as perception's cognitive role is concerned, the distinction between naked-eye perception and instrument mediated perceptions is that the latter involves a longer causal route.

The difference between mediated perception and naked-eye perception is not just causal. To use a telescope to see, the subject needs not only look into one and experience visual sensation, but must know that the visual sensation is informative of an external object. That knowledge requires many other representations, besides the representation of e.g. the moon the subject has when looking into a telescope. Mediated perception must be grounded in other representations. It cannot figure into grounding interactions. Naked-eye perception provides the initial grounding. One may say it is non-derivative or original perception. The non-intentional phenomenological experience enables grounding not as information channel, but as tools which, because they are accessible to the first person, enables the subject to use them in conjunction with action to develop a first person perspective on the world.

5.1.3 Intentional explanations

The subject's perspective sets intentional explanations apart from physical explanations. The two kinds of explanations operate in two different conceptual frameworks. Physical explanations are concerned with the subject's behavior as cause and effect. Intentional

explanations are concerned with the subject's action from the subjects' intentional first person perspective (Moran 2001). They look for the bases, not causes, of the subject's actions (Bennett 1965). The difference in kind shows that they answer two different kinds of question "Why did S do A?" As G. E. M. Anscombe said, intentional explanations are characterized by a special kind of 'why.' The intentional why asks for reasons for action (9). Intentional and physical explanations are not two types of answers to one type of question, but answers to two types of questions, or two theories in two different explanatory frameworks. They are not competitive. Why is there such pressure for us to think they are competitive?

5.1.4 Realism

For many philosophers, our conception of the physical world dominates our conception of what is real. If mental states do not fit into this framework, they are epiphenomenal. If they do, there is a bigger debate about the nature or perhaps the degree of their reality.[20] Our intuitions, or convictions, about the reality of mental states put these states in the physical world, and then we try to understand their nature in terms of their role in the physical world. I think this approach gets it backwards. The reality of anything cannot be evaluated unless we know what that thing is. J. L. Austin calls the concept 'real' "substantive-hungry." "For one and the same object may be both a real x and not a real y; an object looking rather like a duck may be a real decoy duck (not just a toy) but not a real duck" (69). The concept 'real' is like the concept 'good'. We cannot first define the criteria of 'good' and then assess whether something is good or not according to whether it meets those criteria. There is no such of good things. A good book, good weather, a

[20] See Fodor 1990, pp. 3-31.

good idea all have different criteria of what makes each good. These criteria follow from the nature of these things. The concept 'real' behaves the same way. We cannot first decide that what is real is the physical, and then assess whether mental states are real or epiphenomenal according to their role in the physical world. To know whether mental states are real, we need to know what they are, and then determine the criteria of reality according to their nature. If beliefs and desires are essentially part of the intentional framework, then what counts as real in the physical world does not determine the reality of beliefs and desires.

5.1.5 Mind and world

Dretske's question was: how does the mind fit into a world of causes? It seems that science can, in principle, provide complete description of the world, which would make no room for intentional entities. How do we make sense of the idea that intentional states and properties are not epiphenomenal? The problem with this question is the meaning of 'complete'. Given the nature of scientific vocabulary, the progress of science only suggests (at best) that it can provide a complete *physical* description of the world. Intentional explanations utilize concepts which do not provide such descriptions. There is no competition. The problem only arises if by 'complete' we mean all the description we can or want to ask for. In that sense, it is not true that science is able to provide a complete description of the world. In so far as we think intentional explanations are legitimate (which is a premise for the problem), then these are explanations we care for, apart from physical ones.

Pressing on the irreducible distinction between intentional framework and physical framework may lead to a picture of two worlds, one defined by each framework. We ought to

resist such bifurcation, since not only it is unintuitive, it is incoherent. The same objects which are subject to laws of physics are also objects of mental representations. This essay has emphasized how intentionality is grounded in physical interactions. We want to have a coherent world view embodying both physics and intentionality. However the question here is not how intentionality fits into a physical world. Rather we need a larger conception of our world which accommodates both physics and intentionality. I think John McDowell provides such a conception in *Mind and World*. He argues for a broader notion of nature than that which defines science. In my vocabulary, having the intentional potential is natural to certain animals just as having mobility is natural to some organisms, and magnetism is natural to certain metals. "In Aristotle's conception of human beings, rationality is integrally part of their animal nature, and the conception is neither naturalistic in the modern sense (there is no hint of reductiveness or foundationalism) nor fraught with philosophical anxiety" (109).

5.1.6 Structure of content

The intentional structure shows that ideas, and hence, mental contents must be understood in terms of an entire perception-action nexus (which may be enriched by many intermediates). So when we speak of perceptual content, we mean ideas as exercised in perception, not a special kind of idea. The same idea can figure in the subject's belief, and intentions. Modalities do not demarcate one idea from another. Individuation of an idea's content is determined by its perception-action nexus, and its transaction with the world through the two mind-world contact points of perception and action. We should think of individuation and identity of ideas (e.g. a cup vs. a book) much as how we think about the value of money. The purchase value of a bill

does not depend on what the bill looks like, whose hand the bill is currently in. It is determined by what goods (a non-monetary entity) it can be exchanged for. The value of money lies in the interface between the 'money world' and the world of 'real' goods. Its meaning is determined not by the transactions within the money world, but by what non-monetary goods it can purchase.

Content, like the value of a bill, is determined by how it is used in the subject-world transactions. Whether at a particular moment we are talking about the idea as exercised in the subject's belief, or perception, does not identify or individuate the content any more than transactions of a bill between a bank teller and a customer help determine the value of the bill. Like purchasing power, the content of an idea is amorphous and holistic, not attached to any particular modality, but used by all. To regard the idea 'cup' in perception as having different intentional properties than its counterpart in belief or intention is like treating the bill in one bank as having different purchasing power than what it has when it was in another. The intentional structure entails content holism and emphasizes the role of content across different modalities.

5.2 RATIONALITY

5.2.1 Concepts: representations invoking rationality

My theory claims that mental representation can be grounded without invoking reason. In other words, intentionality does not require rationality. A number of philosophers would disagree. If intentionality requires rationality, then there cannot be non-rational animals with mental

representations. This seems counterintuitive. Though we would deny most animals such as dogs and tigers, reason, we would not want to say that they do not genuinely see or want or act. They seem to share mental representation with us, though not rationality. Why should one think that rationality is a prerequisite for intentionality?

One popular approach focuses on the idea that an essential aspect of thoughts and beliefs is their position in a large inferential network of mental states (Davidson 1982). To be rational is to have this framework of an inferential network. To be intentional is to have thoughts. Therefore to attribute thoughts, intentionality, to a system we must attribute rationality. This line of thought is articulated by Davidson, inferentialist theories (Brandom, 1995, 2000), and conceptual role semanticists (Block, 1986). They deny the possibility of non-rational, but intentional animals.

This approach is ineffectual for those who believe in non-rational intentionality. It begins with the observation that of essential features of *our* thoughts, that is thoughts of *rational* animals, rather than an independent notion of intentionality. Their arguments are biased towards rational animals. At most these arguments can show that it is not possible to have thoughts *such as ours*, more generally, intentionality such as ours, without rationality. But we cannot infer from that that there is no non-rational intentionality. Some of their arguments that non-human animals, such as dogs and cats, do not really have beliefs focus on their claim that they do not truly infer. That, again, only shows that they do not have beliefs of our sort, those which are imbedded in a rational framework defined by inferential connections among mental states.

A closely related approach to rationality appeals to the idea that thoughts and desires are attributed as part of making sense of the animal's action. In doing so we must treat the actions as based on beliefs and desires. These mental states necessarily bear logical connections with the

subject's action. To attribute these states one must attribute them as part of a rational framework.

Beliefs and desires certainly make actions sensible, typically to the third person. But the functions of beliefs and desires are to provide bases for action, independent of whether anyone is using this basis relation to make sense of the subject's action. The 'making sense' approach can be a helpful heuristic device to highlight the basis relation. But if it is used as a defining criterion of belief/desire attribution, then the resulting theory is interpreter-centered, rather than subject-centered (Davidson, Dennett 1987). It is too externalist. The theorist may be the one trying to make sense, but it is making sense of the subject from the subject's point of view (here, using the method as a heuristic device). How the subject views the world depends on whether the subject is rational or not. We cannot try to elucidate attribution of rationality by appealing to the subject's perspective.

In order to argue that intentionality is possible with or without rationality, one must argue that the intentional framework is possible with or without the rational framework. We may say that any intentional system's mental states are connected by logical relations, the basis relation mentioned before being one of them. But that is not enough to make inferences. To be able to infer, one must be aware of or better sensitive to, the logical relations among beliefs. She must have a sense that certain propositions provide justification for others. She must have a sense of beliefs as having justificatory status. That sensibility, I suggest, is what distinguishes rational from non-rational animals. I will refer to that sensibility as epistemic sensibility.

Beliefs are not the only mental states which have justificatory status. It also applies to perception and action. Both non-rational and rational animals see. But a rational animal can wonder if what he sees is real or whether his mind is playing tricks on him. He is sensitive to the justificatory status of his seeing. Likewise, we can ask whether our desire is good or bad,

whether a belief ought to be adopted or not.

The epistemic sensibility is not only exercised when we question the justificatory status of a mental state. We use it when we take a justificatory stance. For example, part of the adoption of a belief is taking the stance that the subject believes she *ought* to accept that belief. That is taking an epistemic stance. This attitude is revealed when someone challenges her in saying "P is false" and she replies "Why?" We can only seek for justification of a belief, if we have a sense that beliefs have justificatory statuses.

An animal can represent the world through her desire and action, without having a sense that her desire and the belief using those representations have a justificatory status. An animal has beliefs and acts. But it cannot stand back and question whether this belief is well-founded. It does not have any sense that there are justificatory grounds for beliefs.

5.2.2 Rational and intentional frameworks

Being rational is not a result of having a 'higher' cognitive faculty in addition to the 'lower' faculties such as perception. It is a matter of having epistemic sensibility, which accompanies exercises of all our mental capacities, from perception to action. Rationality is *a way* for the subject to execute her intentional capacities, i.e. her capacity to see, believe, desire, act, etc. It permeates throughout the intentional framework. The difference between a non-rational animal and a rational animal is not that the former only has 'lower level' faculties such as perception, and the latter has, in addition, 'higher level' faculties such as reason. That distinction does not entail such bifurcation of cognitive faculties. The rational framework is the intentional framework imbued with epistemic sensibility.

If this is on the right track, then we have a way out of the conceptual/nonconceptual debate. One strong non-conceptualists argument is the claim that we share representational states, such as perception, with non-rational animals (Peacocke 2001a, 2001b, Evans 1982). On my view, we do share a certain aspect of our mental intentionality with non-rational animals: both rational and non-rational minds are structured in an intentional framework. However, I also agree with McDowell that in the case of our minds we cannot cut off our 'lower states' such as perception from the conceptual domain, for then there cannot be rational links between perception and thinking or judging. Our mental content is conceptual all the way down.

5.2.3 Grounding rational intentionality

This essay has been concerned with grounding intentionality in general, not the particular case of rational intentionality. With the above picture of rational intentionality, we can extend that general structure to conceptual contents. The intentional subject-world link is established by the subject's acting on her perceptual state. Once she makes a commitment in her action, the feedback from the world provides meaning to her perceptual states. Rational mental states acquire meaning the same way, except the commitment is made based on the subject's epistemic sense. A rational subject's adoption of a belief, or making a commitment, is governed by her sense of what is justified.

5.3 CONCLUSION

In this essay I have argued that intentional content is grounded in subject-world interactions in the structure of intentional framework. I have discussed some of the ways this structure can be expanded. There are other applications of this framework that I do not have the space to address. For example, the association of basic perceptions can result in perception of an object as a symbol of another. This provides a basis for understanding linguistic meaning (which however is complicated by the role of reason and social relations).

The purpose driven approach to concepts coheres with our intuitions. Many classes of concepts can be understood, indeed, need to be understood, in terms of the context of certain purposes or desires. For example, social relations such as "friends", "family", "mother", "stranger", and “teacher" arise from our need to relate to others in certain social contexts. Artifact concepts are understood in terms of what the users want from them. Even many abstract concepts are understood in terms of the explanatory needs of their theories. Many of our concepts are shaped by and grounded in our desires. As our desires change, so do our concepts. For example, the idea of the sun has evolved from antiquity when people needed to see it as a god to the modern day where it fits into a largely un-anthropomorphic view of the universe. That view grew from scientists' continual desire to explain their cosmic data. Just as concerns and purposes change through time, they also differ from society to society. Such differences are responsible for different world views. Where different cultures' concerns and purposes do not overlap, we find untranslatable concepts. These cultural differences and changes show how many our concepts, and the world views constituted by them, are driven by our goals and

purposes, whether they be individual or collective. Our experiences, desires, world views and stock of concepts all form an organic whole.

BIBLIOGRAPHY

Anscombe, G. E. M., 1963. *Intentions*. Harvard University Press: Cambridge, MA.

Aristotle. 2002. *De Anima Book II and III*. tr. By E. W. Hamlyn, Oxford University Press: New York.

Austin, J. L. 1964. *Sense and Sensibilia*. Oxford University Press: New York

Ayer, A. J., 1956. *The Problem of Knowledge*. London: Macmillan.

Bennett, D. 1965. "Action, Reason, and Purpose" *The Journal of Philosophy*. vol. 62, No. 4, (Feb. 18, 1965), pp. 85-96.

Block, Ned, 1986, "Advertisement for a Semantics for Psychology," *Midwest Studies in Philosophy*, 10: 615–678. Reprinted in Stephen P. Stich and Ted A. Warfield, eds., *Mental Representation: A Reader*. Oxford: Blackwell, 1994.

Brandom, Robert, 1995, *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.

——, 2000, *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.

-----, 2002, "Non-inferential Knowledge, Perceptual Experience, and Secondary Qualities: Placing McDowell's Empiricism," in N.H. Smith, ed., *Reading McDowell: On Mind and World*. London: Routledge.

Bermúdez, José Luis, 2000. "Naturalized Sense Data," *Philosophy and Phenomenological Research*, 61: 353–74.

Davidson, D. 1985. Incoherence and Irrationality. *Problems of Rationality*. 2004. Oxford: Clarendon House.

----- 1982. Rational Animals. *Subjectivity, Intersubjectivity, Objective*. 2001. Oxford: Clarendon House.

---- 1975, "Thought and Talk," in S. Guttenplan, ed., *Mind and Language*. Oxford: Clarendon Press: 7-23.

Dancy, Jonathan, 1995. “Arguments from Illusion,” *Philosophical Quarterly*. 45: 421–38.

Dennett, D. 1987. *The Intentional Stance*. Cambridge, Mass.: MIT Press.

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

Evans, Gareth. 1982. *The Varieties of Reference*. ed. By John McDowell, Oxford University.

Field, H., 1977. "Probabilistic Semantics". *Journal of Philosophy*. 1977, 74, 379-409.

Fodor, J. A. 1980. "Methodological Solipsism as a Research Strategy in Psychology". *Behavioral and Brain Sciences*. 3, 63-73.

--- 1990. *A Theory of Content and Other Essays*. MIT: Cambridge, MA.

Grice, P. 1957, ‘Meaning’, *The Philosophical Review*, 66: 377-88. Reprinted in SWW.

Gunther, H. Y., 2001. Content, illusion, partition. *Philosophical Studies*, 102: 185–202

Harman, G. 1987, "(Non-Solipsistic) Conceptual Role Semantics," in E. Lepore, ed., *New Directions in Semantics*. London: Academic Press: 55-81.

Jackson, Frank, 1977. *Perception: A Representative Theory*. Cambridge: Cambridge University Press.

Loar, B. 1990 “Phenomenal States” *Philosophical Perspectives 4, Action Theory and Philosophy of Mind*.

----- 1981. *Mind and Meaning*. Cambridge: Cambridge University Press.

McDowell, J., 1994. *Mind and World*, Cambridge MA: Harvard University Press.

Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press: Princeton.

Nagel, T. 1973. What Is It Like to Be a Bat. *Philosophical Review* 83 (October):435-50.

O'Shaughnessy, Brian, 2003. Sense Data. *John Searle*, ed. Barry Smith. Cambridge: Cambridge University Press.

Peacocke, C, 2001a. Does perception have a nonconceptual content? *Journal of Philosophy*, 98: 239–264.

——, 2001b. Phenomenology and nonconceptual content. *Philosophy and Phenomenological Research*, 62(3): 609–615.

Putnam, H. 1974. “The meaning of ‘meaning’,” in H. Putnam, *Philosophical Papers*, vol. II, *Language, Mind and Reality*, Cambridge: Cambridge University Press.

Robinson, Howard, 1994. *Perception*, London: Routledge.

Russell, Bertrand, 1997. *The Problems of Philosophy*, New York: Oxford University Press.