

Natural Language as a Language of Thought

by

J. Alden Jerrett-Enns
B.A., University of Winnipeg, 2005

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF ARTS

in the Department of Philosophy

© J. Alden Jerrett-Enns, 2007
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Natural Language as a Language of Thought

by

J. Alden Jerrett-Enns
B.A., University of Winnipeg, 2005

Supervisory Committee

Dr. James O. Young, (Department of Philosophy)
Supervisor

Dr. Patrick Rysiew, (Department of Philosophy)
Department Member

Dr. Karen Brown, (Department of Philosophy)
Department Member

Dr. Suzanne Urbanczyk, (Department of Linguistics)
External Member

Supervisory Committee

Dr. James O. Young, (Department of Philosophy)
Supervisor

Dr. Patrick Rysiew, (Department of Philosophy)
Department Member

Dr. Karen Brown, (Department of Philosophy)
Department Member

Dr. Suzanne Urbanczyk, (Department of Linguistics)
External Member

Abstract

There is an important debate in the philosophy of mind about whether we think in natural language or a non-conscious innate mental language called mentalese. The latter view relegates natural language to a role as just a tool of communication, something that can offer no insight into the operation of the mind. The aim of my thesis is twofold. First, I wish to show that the arguments in favour of the mentalese hypothesis are wanting and that many of the phenomena that it is supposed to address can be adequately explained by what I will call ‘the natural language as a language of thought hypothesis.’ Second, I will argue that there is strong introspective evidence in support of the natural language as the language of thought hypothesis, at least enough to show that we conduct some of our thinking in natural language. The evidence for and the simplicity of the natural language as a language of thought hypothesis offer good reasons to further investigate the role of natural language in cognition, despite the arguments that Fodor provides to the contrary.

Table of Contents

Title Page.....	i
Supervisory Committee.....	ii
Abstract.....	iii
Table of Contents.....	iv
Dedication.....	v
Introduction.....	1
Chapter 1.....	8
Chapter 2.....	37
Chapter 3.....	60
Conclusion:.....	80
Bibliography.....	81

To Joanne Sefton

Introduction

The responses to the question of the relationship between natural language and thought can generally be divided into two approaches. I will follow Carruthers (1996) in labeling them communicative and cognitive. The debate between these approaches is about whether natural language serves as a medium for thought or just as a tool for its expression. The communicativist believes that the function of natural language is to communicate thoughts. One important communicativist position, defended by philosophers such as Fodor (1975, 1987, 1990) and Pinker (1994), is that we use natural language to articulate thoughts encoded in a language of thought called 'mentalese.'¹ The operation of mentalese is not introspectively accessible. Hence it is hypothesized to exist beneath conscious awareness, at a level of functional organization more sophisticated than the bare neural events of the brain. The cognitivist, on the other hand, believes that natural language performs both a communicative and cognitive function. The cognitivist argues that natural language is the medium in which we conduct our conscious propositional thought. Consequently, the acquisition of natural language is supposed to have an important role in the development and augmentation of our thought processes.

The objective of my thesis is to demonstrate the superiority of the cognitivist understanding of the relationship between natural language and thought. The arguments for the communicative approach are inferior. The communicativist argues that there are phenomena for which the mentalese hypothesis provides the best

¹ To avoid confusion, I will refer to what Fodor calls 'the language of thought' and 'mentalese,' as simply 'mentalese' (I will also change what he calls the 'language of thought hypothesis' to 'the mentalese hypothesis.' The phrase 'the language of thought' I will use to describe the language in which we think, be it mentalese or natural language.

explanation. I will argue that this is mistaken, for the best explanation is actually offered by the cognitive approach. It explains the phenomena as well as (if not better than) the communicative approach, and is supported by strong introspective evidence. Furthermore, the cognitive approach is more parsimonious and has greater potential for fostering scientific investigation into the operation of the mind/brain. The mentalese hypothesis postulates that natural language and thought are completely independent, in spite of the phenomenological feeling that they are often intertwined. Thus, mentalese ignores the potential insight that natural language may offer into the structure and operation of the mind. Supporters of the mentalese hypothesis attempt the daunting task of searching for mentalese in the chaos of our brain activity with few clues to guide them. In contrast, the cognitive approach postulates that the linguistic structure of the brain reflects the natural language monologue that fills our consciousness every day of our lives. This offers some guidance in the study of the mind/brain because we may look for analogues of natural language sentences represented in our neural activity. In addition, it is more promising than the mentalese hypothesis, since we may already have a general idea of where to look for natural language in the operation of the mind, namely in the subsystems of the brain responsible for speech production and comprehension (Hauser 1995).

In spite of the superiority of the cognitive approach in these areas, the mentalese hypothesis receives strong support because it offers a plausible theory of thinking, language acquisition, and the intelligent behaviour of animals and infants. This will be the focus of Chapter One, in which I will discuss and reject these grounds for hypothesizing the existence of mentalese. I will argue that natural

language can provide an explanation that is on better abductive footing, given that the role of natural language in our thinking is introspectively accessible and the fact that it can do the same explanatory work. The operation of mentalese in our thinking has yet to be discovered and this constitutes a strong reason for prioritizing scientific investigation of how natural language structures our thinking. The introspective argument for natural language as a language of thought will be the subject of Chapter Two, in which I will suggest that introspection gives us good reason to believe that at the very least, *conscious*² *propositional* thinking occurs in the medium of natural language.

I will conclude in Chapter Three with an examination of Fodor's semantic argument for the mentalese hypothesis from the assumption that intentional realism is true. Intentional realists believe that the propositional attitudes we use to explain behaviour in terms of beliefs and desires refer to actual states physically realized in the mind/brain. I share Fodor's belief that intentional realism is true, and I will take this for granted as a constraint in determining an acceptable theory of the relationship between natural language and thought. However, Fodor believes that the only theory of meaning that accommodates intentional realism requires that we postulate the existence of mentalese. I will argue that this is not the case by proposing an alternative theory of meaning that not only respects the truth of intentional realism, but also supports the role of natural language as a medium for thought and does not require we postulate the existence mentalese. Before I begin, I would like to make

² We are introspectively aware of conscious thought. Conscious thought occurs at the *personal level* of awareness. This is in contrast to non-conscious thought, which occurs without awareness or intention, like in cases visual image processing. Thought of this type occurs at the *subpersonal level*.

some preliminary remarks about what mentalese is supposed to be and the general motivations for accepting its hypothesis.

The mentalese hypothesis can be divided into three fundamental claims (Bermudez 2003: 23). The first is that intentional realism is true. This means we can, and eventually will, succeed in naturalizing the mind through the discovery of the physical instantiation in the brain of the everyday propositional attitudes we use to explain our own behaviour and the behaviour of others. Not only does this offer hope for scientific vindication of common sense folk psychology, which has been a powerful tool in predicting and explaining human behaviour for thousands of years, but it also offers a very plausible theory of thinking. The mentalese hypothesis shows how the inferential transitions between thoughts can be realized in physical structures, and how these physical structures, as the physical realizations of propositional attitudes, can be causes of behaviour.

Second, the mentalese hypothesis supposes that these physical structures are sentential. This means that these physical structures are logically isomorphic with the propositional attitudes they represent. As representations of propositional attitudes, these physical structures also reflect the compositional structure of propositional attitudes. And as propositional attitudes are divisible into semantic components that may be recombined according to syntactic rules, so are sentences of mentalese.

Third, the mentalese hypothesis claims that the causal transitions between the physical states realized in the brain respect the rational relations between the thoughts those states represent in the mind. The mentalese hypothesis provides a physical explanation of how beliefs and desires combine to cause intentions and how thoughts

interact to produce new beliefs and new desires. Of course, someone might wonder how this is possible.

In *Psychosemantics*, Fodor argues that if we assume that having a propositional attitude entails the tokening of a symbol in mentalese, then we may explain how the causal properties of a thought can yield its syntax. This, in turn, will provide the key to its semantics. The idea is that the syntax of a symbol determines the causes and effects of its tokenings in the same way that the geometry of a key determines the locks it will open (Fodor 1987: 19). This syntactic shape prevents the organization of semantic elements to produce nonsense, guaranteeing that we do not end up with sentences — hence thoughts — like, ‘There are four prime numbers in Caesar’ that are formed according to syntactic rules but produce semantic nonsense. The syntactic shape of a symbol also promises to ensure that it respects logical transitions, so computations that are performed on a symbol adhere to the basic tenets of good reasoning. This explains why we do not find minds that can apply basic logical transitions like conjunction or *modus ponens* in one instance but not in others.

The mentalese hypothesis provides a model for thinking that is demonstrably successful in the operation of digital computers. And it shows how the mind could operate analogously to such a computer. The software of the mind is supposed to be physically instantiated on the hardware of the brain, in the same way that the ones and zeroes hidden within my computer physically represent the words on my computer screen. The sentences on my computer have nothing to do with the ones and zeroes that represent them; the mentalese hypothesis supposes that the relationship between

the sentences of natural language present to the mind and the sentences of mentalese is analogous.

The problem is that I cannot get to the physical instantiation of the sentences on my computer screen without opening it up and looking inside. The mentalese hypothesis posits that the relationship is the same between mentalese and the sentences of natural language present to the mind. For this reason mentalese is postulated to exist at the subpersonal level, beyond the reach of our conscious awareness. We may object to the mentalese hypothesis by suggesting that natural language goes all the way down in structuring the way in which we think. There are good reasons to suppose that natural language serves as a linguistic vehicle for the representation of conscious propositional thought and the computations performed upon it, as opposed to the introspectively inaccessible and empirically unverified language of mentalese.

The best argument for the existence of the mentalese is that certain types of thought are not be possible without linguistic vehicles to provide a medium in which they can be represented (Bermudez 2003: 26), and that natural language cannot serve as such a medium. Fodor rules out natural language as a linguistic vehicle for thought due to the many examples of thought occurring in the absence of natural language. The examples include language acquisition, the intelligent behaviour of animals, and the intelligent behaviour of infants. All of these thought processes require a medium of representation, and the medium cannot be offered by natural language, as none of these organisms can use natural language. Fodor concludes that organisms that lack natural language must therefore represent their environment in mentalese. This much

may be true. However, Fodor goes on to claim that because natural language cannot be a medium of representation in some cases, it cannot serve as a medium of representation in all cases. The inference from the existence of thinking organisms that lack natural language to the impossibility of natural language as a language of thought is dubious. If this can be shown, the explanatory power of mentalese will be greatly diminished and much of its appeal as a hypothesis defeated. This is what I will attempt in my opening chapter.

Chapter 1:

The Virtues of the Natural Language as a Language of Thought Hypothesis

In this chapter I will canvass what I believe to be the strongest arguments for the mentalese hypothesis. I will challenge Fodor's claim that postulating the existence of an innate combinatorial system of symbolic representations (referred to as mentalese, for brevity's sake) provides the *best* explanation of phenomena such as the structure of natural language, language acquisition, and the intelligent behaviour of nonlinguistic animals and prelinguistic humans. Evidence for the intelligent behaviour of animals and infants motivates Fodor's flippant dismissal of the proposal that the medium of some adult human thought is natural language (Fodor 1975: 56). The thesis of this chapter is that the evidence is inexcusably scant to support such a conclusion, and, in fact, there is actually strong evidence to the contrary. There may not be a decisive argument for the conclusion that natural language sentences are the only vehicles for our thoughts. However, there are arguments that show that the proposition that natural language is constitutive of thought is at least on par with the mentalese hypothesis. In the case of a draw, I believe that we should favour natural language as the medium in which we consciously represent our environment. We should favour this supposition as more plausible than the mentalese hypothesis for two reasons. First, it does not require mentalese's potential commitments to nativism. Second, it is more parsimonious because it needs only one introspectively accessible symbolic system to explain human thought, as opposed to the introspectively inaccessible and arguably undiscoverable language of mentalese.

The structure of this chapter is as follows. First, I will summarize and repudiate Fodor's arguments for the existence of mentalese on the basis of linguistic structure, since supposing that natural language is a medium in which we conduct our conscious propositional thought explains these phenomena just as well as mentalese. Second, I will show that many of the empirical issues — in particular, aphasia, feral children, and the historical figures Albert Einstein and Helen Keller — pertinent to the mentalese-versus-natural-language-debate fail to offer decisive evidence for either side. Third, I will take issue with Fodor's claim that there is indisputable evidence that natural language cannot serve as a medium of representation because there is evidence of thought occurring in the absence of natural language in animals and infants. Fourth, I will discuss the potential epistemic problems with the mentalese hypothesis. In particular, scientific investigation has failed to find mentalese operating within the brain and our introspection of the mind reveals only natural language integrated with various images. These four sections combine to demonstrate that we should explain thought in terms of natural language, because it offers a more ontologically parsimonious explanation of the phenomena that are supposed to motivate the mentalese hypothesis. The hypothesis that natural language is medium in which we conduct our conscious propositional thought requires one representational system instead of two. As such, if we accept that a simpler theory is preferable to a complex one (provided that it receives as much if not greater evidential support), then the natural language as a language of thought hypothesis is preferable as a simpler and potentially more fruitful theory for further scientific investigation.

There are a variety of reasons to prefer simpler theories. First, there is the pragmatic value that simpler theories are easier to work with. Second, it is a common sense methodological principle to favour the simplest theory that adequately explains the phenomena without postulating additional unnecessary entities. It provides a starting point in our theorizing that is not only makes it easier to manage but historically proven to be successful. Third, there is the fact that falsifiability is a virtue of scientific investigation, and simpler theories offer greater opportunity to test their veracity by postulating fewer entities for verification. Popper clearly sums up this point by saying that “[simpler theories are preferable] because their empirical content is greater; and because they are better testable” (Popper 1992). In this debate, there is strong evidence on the side of the natural language as a language of thought hypothesis in the form of introspective awareness that natural language plays an important cognitive role as a representational medium for our thoughts.

1. Linguistic Phenomena

Productivity

Thought is productive in the sense that we are capable of understanding a potentially infinite number of thoughts, even though we entertain only a finite number. For example, if someone can think that John has father, then she can also think that John’s father has a father, John’s father’s father has a father, and so on to infinity. Furthermore, there are an infinite number of sentences that we have never heard before and yet would still be able to understand. Likely at least one such sentence is written in this thesis. The challenge is coming up with a good explanation for the fact

that we have a capacity to understand an infinite number of thoughts, even though the number we are able to entertain is finite. We are finite beings and so we have neither the time to entertain nor the storage capacity to contain an infinite number of mental representations. So how do we get an infinite capacity for understanding from finite mental resources?

Fodor and Peter Carruthers³ share the belief that the explanation lies in the appeal to the recursive structure of the syntax of a language physically instantiated in the brain. Moreover, positing the existence of such a language has the twofold advantage of explaining the link between meaning and mind and showing how beliefs and desires can be as causally efficacious as common sense psychology takes them to be. First, assuming that the mind conducts computations on mental representations with respect to their syntax, we can understand the meaning of these representations because the rules for how representations are put together must respect their semantic properties. Second, if we accept that tokenings of representations are physical particulars, then there can be no problem with explaining how they have causal powers. Postulating the existence of a mental language, either the innate language of mentalese or supposing that natural language, once acquired, becomes constitutive of conscious propositional thought, not only explains productivity, but also provides an account of the mind that coheres with common sense psychology and intentionality.

The argument for the existence of some linguistic structure of mind on the basis of productivity goes like this:

³ He is one of the most prominent proponents of the natural language as a language of thought hypothesis. His position is best articulated in *Language, thought and consciousness* (1996).

1. The best explanation for productivity is by appeal to the recursive syntax of a language.
2. Natural language is productive if and only if thought is productive, because we can think all the sentences we can say and say all the sentences we can think.
3. Natural language is productive.
4. Thought is productive. (2, 3)
5. If thought is productive, then it is either productive as a result of the acquisition of natural language and its recursive syntax, or the recursive syntax of mentalese causes the productivity of thought.

(C) Therefore, either natural language and its recursive syntax are responsible for the productivity of thought, or the productivity of thought is owed to the recursive syntax of mentalese. (4, 5)

The productivity argument is not an argument for the mentalese hypothesis as much as it is an argument for the linguistic structure of thought. We may leave this argument, then, with the conclusion that it has nothing important to say about the language in which we think, only that it motivates the supposition that we think in some kind of language, be it natural language or mentalese. Fodor does not put much stock in this argument because those who oppose understanding thought linguistically can simply deny (1) (1987: 148). The argument requires that we idealize beyond our ability to understand a finite number of sentences over a finite period of time and extrapolate to the assumption that we can understand an infinite number of sentences. So an objector may refuse to idealize and propose an account of how we come to

understand the many sentences we do in dispositional terms, without appealing to the linguistic nature of thought. For this reason, the productivity argument by itself is very weak in establishing that thought is linguistic, and that is why Fodor appeals to systematicity, which does not require any agreement about idealization.

Systematicity

Thought is systematic in the sense that the ability to entertain a particular thought is intrinsically connected to the ability to entertain a variety of other related thoughts. This point becomes clear when we contrast the way we actually learn how to speak our first language with memorizing a gigantic phrasebook. It is possible to learn any part of a phrasebook without learning the rest. For example, I can say 'How are you?' in Korean, ('Annyong haseyo') without understanding anything else in Korean. This is not the case for native speakers of English, since when one understands a sentence one also understands a collection of other systematically related sentences. For example, there are no native English speakers who know how to say 'John loves Mary' who cannot also say that 'Mary loves John.' In other words, there are no cases in which we find a native English speaker who only knows one or two unrelated sentences, like someone learning a language out of a phrasebook. Rather, the competency of a native English speaker is systematic. This fact is as solid as psychological laws come.

The question is how best to explain why systematicity is a psychological law. This is best answered when we consider what takes place when someone acquires a language. The phrasebook explanation is obviously unsatisfactory. The alternative is to explain acquiring a language in terms of mastering its syntactic rules concomitantly

with learning its vocabulary. Once one knows the rules for properly constructing a particular sentence out of a set of words, we can see how it follows that a reapplication of the rules and a reorganization of the words in the set can be carried out numerous times to produce many different sentences. In fact, an infinite number of sentences can be produced if you are willing to idealize. This is possible only if we do not take sentences to be the atomic unit of meaning. If we treat sentences as the smallest unit of meaning, we have no recourse to explain systematicity in terms of the recombination of semantic elements. Instead, we must suppose that sentences are compositional, that is, the meaning of a sentence is the product of its component parts in tandem with its syntactic structure. The upshot of this is that we have to endorse a combinatorial semantics in addition to a combinatorial syntax. Fodor explains that the reason for this is straightforward. "You can't construct the meaning of an object out of the meanings of its constituents unless it *has* constituents" (1987: 150).

It is the combinatorial semantics of language that allows for the recombination of words in a particular sentences according to syntactic rules in order to produce many semantically different, yet systematically related sentences. These two sentences contain the same words but have different meanings because the semantic elements are reorganized according to the syntactic rules of the language: 'John loves Mary and Mary loves John and the fish loves no one' or I can say that 'Mary loves Mary and John loves no one and the fish loves the fish.' At the same time, this explains why the ability to understand certain sentences is systematically related to the ability to understand other sentences that have lexical elements in common,

whereas there is no connection in understanding between the sentences 'The cow jumped over the moon' and 'Nine is a prime number.'

The inference from the systematicity of language to the systematicity of thought is not a very bold step. But, as in the case of productivity, it appears that we may just as easily suppose that natural language is responsible for the systematicity of thought, instead of accepting Fodor's conclusion that the systematicity of thought produces the systematicity of natural language. Regardless, it is clear that if language is systematic then so is thought and vice-versa. Just as it is a psychological law that the ability to understand a particular sentence is linked to the ability to understand many others, it is a psychological law that the ability to think a particular thought is linked to the ability to think many others. This is because being able to understand a sentence is simply being able to think the thought or proposition that it expresses (Aydede 2004). However, Fodor thinks that he has license to infer more than this from the systematicity of thought. He thinks that systematicity somehow establishes the existence of mentalese. The following is my summary of his argument:

1. If thoughts have constituent structure, then mentalese exists.
2. Linguistic capacities are systematic if and only if cognitive capacities are systematic, since the ability to understand a particular sentence and the sentences systematically related to it is identical with the ability to think a particular thought and the thoughts systematically related to it. Thus, cognitive capacities must be at least as systematic as linguistic capacities.
3. Linguistic capacities are systematic because sentences have constituent structure.

4. Cognitive capacities are systematic for the same reason that linguistic capacities are (on the basis of (2)), because thoughts have constituent structure.

(C) Therefore, mentalese must exist (Fodor 1987: 150-151).

The problem with this argument is that (1) is incorrectly formulated. If thoughts have a constituent structure, then mentalese *may* exist, but we can just as easily explain the constituent structure of thought by supposing that natural language is responsible for organizing thought such that it has constituent structure. As in the case of productivity above, it all depends on whether one is willing to give explanatory priority to thought or natural language. In the opening paragraph of this chapter, I argued that in the case that we could not arbitrate a clear winner of the mentalese-versus-natural-language-debate, we should favour natural language. This is because it offers the simplest and most easily accessible explanation of linguistic phenomena, as we are not even consciously aware of the operation of mentalese.

Since the operation of mentalese is not introspectively accessible it is hypothesized to exist beneath conscious awareness, but at a level of functional organization more sophisticated than the bare neural events of the brain. It will remain a mere hypothesis until it receives scientific vindication, but scientific investigation into the existence of mentalese has been so fruitless thus far that some critics of the hypothesis have gone so far as to call it beleaguered (Hauser 1995). It is my contention that the study of the relationship between language and thought should be reoriented away from the communicative model. Instead, the attention of scientists should be refocused on the study of the role of natural language in our cognition.

There are three reasons for this. First, we can see the significance of natural language in cognitive development of infants. Second, there is colloquial and scientific evidence that shows that a child's ability to think is closely related to their ability to use natural language (see Carruthers' 1996: 40-47). Third, we experience the role natural language plays in our cognition introspectively as adults when we feel that our thought occurs in the form of inner speech, heard in the auditory imagination.

My argument so far is strikingly similar to the thesis of Brent Silby. He concisely summarizes the strategy I have taken up to this point as follows:

1. Human thinking exhibits productivity and systematicity.
 2. Productive and systematic thoughts require syntactic structure.
 3. The required syntactic structure can be explained by either mentalese or natural language.
 4. We should favour economy in our ontology (this is to say: we should avoid positing the existence of something that we cannot locate in the empirical world if the phenomenon in question can be explained by something that we can locate in the empirical world).
 5. We cannot locate mentalese or the brain mechanism responsible for mentalese.
 6. We can locate natural language and there is good evidence to support its role in thinking.
- (C1) Natural language should be used to explain the syntactic manipulation required for producing productive and systematic thinking.
- (C2) Therefore, our thoughts should be explained by our use of natural language (Silby 2000).

This is Silby's argument that human thinking is best understood in terms of our usage of natural language. Independently of reading his thesis, I arrived at a similar argument, which I stated at the beginning of this chapter, and have restated in this section's discussion of the arguments from linguistic phenomena to the existence of mentalese. At most, the linguistic phenomena establish that thought is linguistic, but they do not establish whether the language responsible for the linguistic nature of thought is mentalese or natural language.

Fodor does not believe that the linguistic phenomena result in the abductive disjunction described in (3) because he believes that there is strong evidence in the form of the intelligent behaviour of animals and infants that rules out the claim that

the language in which we think is natural language (Fodor 1975). Fodor believes that this proves that there is thought independent of language, so it is not possible that natural language is constitutive of thought. The major flaw in Silby's thesis is that he fails to acknowledge Fodor's rebuttal of (3). The main purpose of this chapter is to answer Fodor's criticism of the possibility that natural language is the language in which we think and thereby make (3) a plausible claim.

I propose to criticize Fodor's position empirically and conceptually. Firstly, there is a significant amount of empirical evidence that has potential to offer insight into the debate between natural language plays a cognitive role in thought versus simply communicating it. There are cases of aphasia, in which brain damage severely limits linguistic capabilities, and cases of abnormal development in which for a variety of reasons individuals lacked the resources to develop natural language. I will argue that the balance of the evidence offers very little that conclusively favours either side. Secondly, Fodor invalidly criticizes the thesis that natural language is constitutive of some thought on the grounds that there are nonlinguistic animals and prelinguistic infants that display intelligent behaviour consistent with what we understand to be thinking. In response I will counter that it does not follow that what applies for animals should also apply for human beings. As for instances of intelligent behaviour among infants, we should consider the possibility that human beings employ different types of thinking, and that the type of thinking displayed by nonlinguistic infants is not the kind for which natural language is the medium. I will conclude the chapter with a defense of premises (4)-(6), discussing the

epistemological and ontological reasons why we should favour natural language as a medium in which we conduct our conscious propositional thought.

2. Empirical Issues

Aphasia

Aphasiacs are people who have lost differing levels of linguistic ability as a result of head injury or stroke. There are different types of aphasia; some only affect speech output, either by eliminating it or rendering it meaningless gibberish, whereas others retard comprehension, word-memory, and grammatical ability. The severity of the impairment suffered by an aphasiac is generally measured in terms of differing levels of speech output rather than comprehension. The reason for this is obvious, given that aphasiacs who can understand instructions make much better test subjects. However, aphasia that limits speech-production is not interesting to the debate at hand. There seems to be no good evidence against the possibility that inner speech is still produced in the auditory imagination of speech-production limited aphasiacs after injury or stroke impedes their speech production ability, even if it may not be the subject of vocalization (Carruthers 1996: 47). For this reason, I propose to set this type of aphasia aside.

As far as the debate about the role of natural language in thought is concerned global aphasia is the most interesting sort. For these people both speech production and comprehension is extremely limited. However, studies conducted with global aphasiacs are problematic because someone has yet to come up with a nonlinguistic test of cognitive ability. Raven's Matrices, which bear some resemblance to jigsaw

puzzles, have been used to study the cognitive impairments of global aphasiacs, and the results have shown a degree of impairment in these nonverbal tasks consonant with the subject's level of language-comprehension impairment. But as promising as this evidence seems as support for the cognitivist understanding of language and thought, it is also in keeping with the communicative picture of language being used simply to express sentences encoded in mentalese. This is because the brain damage in global aphasiacs is so extensive that it may be the case that damage to the areas of the brain responsible for language production and comprehension is also in the same region of the brain that is responsible for cognitive processes that may or may not be encoded in mentalese. Thus, Carruthers supposes that "it might then be that any brain-injury which damages one will necessarily damage the other, without the one actually involving the other" (Carruthers 1996: 48). So for this reason it is best to be cautious in drawing conclusions about the relationship between language and thought from people who have suffered brain damage.

Feral Children

Feral children are human children who are raised in total isolation from human contact at a very young age, either brought up in the wild by animals or confined to solitude by their parents. The fact that feral children are cited both by those who are opposed to (Abbott 1995) and in favour of (Hauser 1995) the cognitive conception of the relationship between language and thought should be telling that such empirical cases are not going to do anything but reconfirm antecedent commitments. Hauser supposes that if the mentalese hypothesis were true, a feral child should display behaviour of clearly productive thought that would be inexplicable without resorting

to appealing to the recursive syntax and compositional semantics of mentalese. But he says that it is highly dubious that a language-deprived human could produce such behaviour on par with that of fully competent speaker of natural language, and thus counts feral children as evidence against the mentalese hypothesis.

On the other hand, Abbott cites examples from the feral child Genie's communication, drawn from Curtiss's research (1988: 98), as evidence that she possessed some form of recursive syntax in mentalese that clearly did not manifest itself in her English. Curtiss provides the translations in brackets below.

- a. Genie bad cold live father house. ('I had a bad cold when I lived in my father's house.')
 - b. Father hit Genie cry long time ago. ('When my father hit me, I cried, a long time ago.')
 - c. Genie have Mama have baby grow up ('I have a Mama who has a baby who grew up.')
- (Curtiss 1988: 98).

These are examples of memories that Genie recounted of a time before she had language. It is questionable whether these memories are evidence of thought without natural language ability or of a reinterpretation of her experience once she had acquired some rudimentary form of English. It is also questionable whether we could really count these sentences as recursive, and if so, whether we should credit their recursiveness to mentalese or the slow development of basic English syntax.

The problem with the case of Genie in particular and the cases feral children in general is twofold. First, feral children are generally denied human contact, which is part and parcel with normal human development. For example, Genie was kept locked up in isolation by her family, held in place with a makeshift straightjacket. She received occasional visits from her father and brother, who made no attempt to communicate with her, and when she made any sound her father beat her. When Genie was rescued at the age of thirteen she had the mental age of a toddler. After she

acquired rudimentary language skills, some disputed that she displayed intelligence in the normal range for her age (Carruthers 1996: 42), in spite of the fact that her language usage remained abnormal (Curtiss 1977: 42).

Even if we were to suppose that language acquisition played an important role in her mental development after she was discovered, we cannot credit Genie's incredible deficit in cognitive abilities to a lack of language, as she also lacked normal emotional contact with other human beings and the freedom to physically move around in and explore her environment. Cognitivists may suppose that the retardation of Genie's mental development was due to a lack of language. However, those who support the communicative conception of language can rebut this claim with the suggestion that it was her lack of normal human contact that was responsible. If there is a close link between social development and linguistic development, and it is plausible that this may be, Genie's lack of a nurturing social environment may have severely hindered her linguistic development. Second, the communicativist may interpret the lag in the cognitive development of feral children as being the product of lacking natural language only insofar as natural language is a precondition for acquiring the ability to conduct certain types or levels of thought. We are all aware that education is known to increase the variety and sophistication of the thoughts one can have. On the other hand, the cognitivist may suppose that the lack of natural language is responsible for the lag in the cognitive development of feral children, because it means that feral children have no medium in which to entertain various kinds of thoughts. The point is that both the cognitive and the communicative

conception of the relation between language and thought is compatible with the evidence from feral children and it favours neither interpretation.

Einstein's Claim to Thought Without Language

The one thing that Genie and Einstein both have in common is that they claim to have had languageless thoughts at one point or another. The source of the claim in Genie's case is obviously because she lacked the words to think those thoughts at that time in her life. The source of the claim in Einstein's case is simply that he conducted many of his thought experiments in visual images and not words. In addition, there are the cases, which I will discuss in a later chapter, in which test subjects reported having thoughts that were neither encoded in image nor language. These three cases pose a problem to the claim that natural language is a medium in which we think.

Curiously, Carruthers deals with Einstein differently than Genie or the test subjects. Carruthers supposes that the creative thinking in which Einstein was engaged occurred beneath the level of consciousness (Carruthers 1996: 58). Why Carruthers makes this move is puzzling, as Silby points out, because Carruthers could just as easily appeal to his compelling idea that in some cases people may have hybrid-ideas that embed propositions in the context of images (Carruthers 1996: 36). Silby gives the following example of how we might interpret how Einstein thought in hybrid image-propositions: "If I were travelling at the speed of light, objects in front of me would look like this [IMAGE], and objects behind me would look like this [IMAGE]" (Silby 2000).⁴

⁴ Carruthers actual position appears to be out of sync with his earlier claim that such cases in which creative thinking is at play, when composers or artists may be thinking totally in terms of images or

Due to the fact that there exists indisputable empirical evidence of thought without language in the form of people who claim to have thoughts that are expressed in their consciousness wholly in the form of images or sounds and without language, Carruthers limits the scope of his thesis, as I am limiting mine, to the claim that *conscious propositional thought is conducted in natural language*. With this demarcation of the cognitive role of language in mind, we may suggest that Genie and the test subjects may have mistakenly thought they were conscious of thoughts that were in the medium of neither language nor images. Or, we may suppose that through a process of retrospective self-attribution and reinterpretation of experiences that were previously non-conscious before the individuals in question applied language to them. The evidence that we may derive from how Helen Keller describes the cognitive transition from being bereft of language to being a language user supports the latter theory.

Before going any further, a note of caution is necessary. By no means do I suppose that isolated first-person testimony should decide this issue, whether Einstein, Genie, or Helen Keller offers it. However, I believe that these cases of first-person testimony provide insight into different ways that the empirical data may be construed in understanding the role of natural language in cognition, in addition to providing motivation for a refined and nuanced understanding of it. Ultimately, even rigorous and systematic testing of multiple subjects will not decide this issue, which is why I will resort to a conceptual criticism of the rejection of the role of natural language as a medium for thought in the next section.

sounds, such thoughts are best understood as being *non-propositional* but not *non-conscious* (Carruthers 1996: 36).

Helen Keller's Claim to No Thought Without Language

Keller went blind and deaf between the ages of one and two. She remained languageless until the age of seven, when she was able to learn language through touch. She learned language through repetitive exposure to certain objects, after which her teacher would trace the name of the object onto her hand. Eventually she mastered this language of touch and went on to write books about her languageless experience. Helen Keller is perhaps a better case study than Genie because she did not suffer psychological abuse or isolation from human contact, so any lag in her development may more easily be credited to lack of exposure to language rather than a lack of exposure to normal human nurturing. Keller's introspection is no more likely to establish or undermine the cognitive conception of language than Einstein's, or mine, or yours, but her introspection is unique, like Genie's, in that she can remember a time when she did not have language. However, Keller's description of her experience without language appears to vindicate the claim that language is constitutive of conscious propositional thought. The quotes from Keller below are taken from Silby 2000.

Before my teacher came to me, I did not know that I am. I lived in a world that was a no-world. I cannot hope to describe adequately that unconscious, yet conscious time of nothingness. I did not know that I knew aught, or that I lived or acted or desired. I had neither will nor intellect (Keller 1909: 141).

Keller's prelinguistic consciousness was a time of limbo. Before she developed language, she claims that she had no inner life and her behaviour was instinctual like that of animals (Keller 1909: 143). Even more interesting is her claim that she had no power of thought and could not compare one mental state with another, which is consistent with her claim above that "I did not know that I knew

aught.” She had no propositional knowledge and no reflexive knowledge of her own self-consciousness. Keller’s introspective claims offer great strength to the introspective arguments of Bermudez and Carruthers, which contend that the reflexive thinking that linguistic ability enables is constitutive of conscious thought. As we shall see, Bermudez and Carruthers both agree that natural language makes reflexive thinking possible, because it allows us to think about and evaluate our thoughts. Carruthers argues that this reflexive thinking is constitutive of consciousness, which is in keeping with Helen Keller’s claim that she lacked consciousness before she developed linguistic ability. Furthermore, Keller says that she did not have a sense of identity before she possessed language and it was only after she learned language she was able to think, with her inner speech taking the form of words being spelled out on her hand (Keller 1909: 145). In sum, Keller’s description of her lack of conscious propositional thought and of her inability think reflexively or compare her mental states represents one instance in which the cognitive conception of language endorsed by Carruthers rings true. But then again, this is only one case, and as my discussion of other empirical evidence up this point has shown, such cases can be finessed as exceptions to either the cognitive or communicative conception of language. Hence, I will now turn to the conceptual criticism of Fodor’s rejection of the thesis that natural language is a language in which we think because animals and infants display thinking behaviour in the absence of natural language.

3. Animals and Infants

In order to understand Fodor's claim that there are nonlinguistic animals and prelinguistic human beings that think, we first need to understand what Fodor means by thinking. Fodor argues that computational processes are the only plausible psychological models of cognitive processes. Paradigm examples of cognitive processes are considered action and concept learning. Computational processes require a representational system in which to compute. Therefore, we must attribute a representational system to all organisms that we believe to have cognitive processes. Fodor argues that natural language will not do as a representational system as follows: "the obvious (and, I should have thought, sufficient) refutation of the claim that natural languages are the medium of thought is that there are nonverbal organisms that think" (Fodor 1975: 56). Fodor believes that the supposition that natural language is a medium in which we think is impossible to take seriously because there exists evidence of nonlinguistic thought, among prelinguistic humans and nonlinguistic animals, so natural language cannot be constitutive of or necessary for thought. I will examine considered action of animals and concept learning in the context of infant language acquisition in detail in order to determine if either constitutes an example of a cognitive process that warrants attributing a representational system to nonlinguistic organisms, and if this is so, if it discredits the claim that the representational system of adult human beings is natural language.

Considered Action

Fodor describes how considered action is carried out in a representational system as follows (Fodor 1975: 28-29). An organism takes her circumstances to be such-and-so. Next she believes that these circumstances offer her a set of behavioural options that she believes she may carry out. Then she assigns consequences to each option and probabilities to each consequence of the various actions she believes to be at her disposal, assigning a preference to each one. Finally, the agent acts in accordance with the behaviour that she deems most suitable according to the preference and probability assigned. According to this model, making decisions is a computational process, carried out in a representational system. The question is whether nonlinguistic or prelinguistic organisms make decisions in this way. If the answer is no, then we may suppose that natural language serves as a sufficient medium of representation such that human beings are able to make decisions. Under what conditions should the answer be yes?

If indeed nonlinguistic creatures display behaviour consistent with the picture of considered action given above or other cognitive processes, then we will also have to accept Fodor's claim that there are nonlinguistic organisms that think. Folk psychological representations entail attribution of cognitive processes such as instrumental reasoning and considered action, instances of thinking which are best understood as computational processes and thus require a representational medium in which to perform computations. However, this does not mean that the representational system in which human beings conduct all of their thinking is mentalese. Chimpanzees can swing through the trees with the greatest of ease, but

even though I share more than ninety-five percent of my DNA with these primates, it does not follow that I should be able to nimbly traverse jungle branches. Nothing follows about human representational systems from animal representational systems without the demonstration of the universality of representational systems.

The most the evidence of animal cognitive processes shows is that the mentalese hypothesis may be justified as an explanation of nonlinguistic thought, provided that we assume that the common sense folk psychological explanations we use to make sense of our fellow humans should do for animals as well. This lends support to Fodor's argument that human beings represent the world in mentalese insofar as it shows that natural language is not necessary for or constitutive of certain types of thought. It defeats the claim that natural language is constitutive of *all* types of thought, but this is not what anyone is suggesting. Fodor would like to claim that he could get more than that from the admitted truth that animals think. He supposes that differences between human minds and animal minds are quantitative (Fodor 2003: 4) and baldly asserts the premise that what is true of one species' system of representation should be true of another. But he provides no evidence to support this claim and the burden of proof is on him, given the fact that human linguistic representation is obviously distinct from nonlinguistic animal representation (if it exists at all) in that human thought is demonstratively (infinitely) productive, systematic, and self-conscious. To ascribe any of these traits to animal cognition puts us on shaky empirical ground. Thus, we may conclude that whether we answer "yay" or "nay" to the question of animal cognition has no bearing on the thesis that natural language is the representational medium of human cognition.

Language Acquisition

It is more difficult to take the same tack with infants as with animals, since the concession that infants think in mentalese would beg the question of why we should treat human thought as being all that different. After all, as Fodor points out, “babies turn into us” (Fodor 2003: 4).

The strongest argument that Fodor provides for the existence of mentalese is that “we have no notion of how any kind of concept is learned except by hypothesis formation and confirmation” (Fodor 1975: 58). (We may safely assume that the same goes for learning words as learning concepts.) Hypothesis formation and confirmation presupposes the existence of a representational medium in which to form and test hypotheses. This medium cannot be natural language, on pain of vicious circularity, since obviously infants cannot represent in a medium they have not yet acquired. This is where mentalese comes in as the medium that makes possible the cognitive processes necessary to acquire one’s first language. As far as Fodor is concerned, this is our best and only option in explaining how infants acquire language. For a representational medium is obviously required for this sophisticated cognitive process.⁵

⁵ To briefly consider Fodor’s model of language acquisition in detail, Fodor says that our theory of language acquisition must explain how experiences of *x*s as *F*, instead of experience of say, *x*s that are *G*, lead an infant to arrive at the conclusion that all *x*s are *F*. Or in plain terms, our theory has to explain how experiences that confirm the hypothesis that the family pet is a cat, instead of a dog, bring an infant to conclude that all experiences of animals similar to the family pet are cats. Our theory can account for this as follows:

[I]f we assume that (a) the organism represents the relevant experiences as experiences of *x*s which are *F*; (b) that one of the hypotheses that the organism entertains about its environment is the hypothesis that perhaps all *x*s are *F*; and (c) that the organism employs, in the fixation of its beliefs a rule of confirmation that all the observed *x*s being *F* is, *certeteris paribus*, grounds for believing that all *x*s are *F* (Fodor 1975: 37-38).

Carruthers and Bermudez both reply to this argument with a distinction between thinking-how and thinking-that. This is a compelling option since we often talk of acquiring a first language in terms of learning *how* to speak and we generally understand knowing the meaning of a word in terms of being *able* to use it correctly (Carruthers 1996: 67). Suppose we understand language acquisition in the non-propositional terms of thinking-how, then perhaps a linguistic representational system would be no more necessary for learning a language than it is for learning how to ride a bicycle. Children obviously do not form and test hypotheses about the physics necessary for staying balanced on their first two-wheeler (Carruthers 1996: 68). Moreover, there are other instances in which understanding thinking-how as non-propositional thought may prove to be a powerful explanatory tool in understanding such things as the cognitive processes of animals, the reasoning that athletes conduct while performing complex bodily skills, and the imagistic or creative thought of artists and musicians (Bermudez 2003: 35). The distinction between thinking-how and thinking-that is that the latter constitutes propositional thought that is inferential in structure and determinate in content, whereas the former lacks determinate content and is best understood as an activity sensitive to the relation between an individual and the perceptual world rather than as a cognitive process.

Bermudez and Carruthers suppose that thinking-how is a potential way of explaining language acquisition at the personal level, the level of conscious awareness, without resorting to hypothesizing the existence of a second representational system in order to explain how natural language is acquired. The problem is that we do not know what is going on at the subpersonal level of adults,

never mind infants. This leaves Fodor with a simple rejoinder that both Bermudez and Carruthers readily admit. Thinking-how could very well be cashed out beneath the level of conscious awareness by the defender of the mentalese hypothesis.

The result of supposing that language acquisition and other examples of thinking-how are a matter of computations in mentalese brings us to the conclusion that the creative thinking of Einstein, artists, and the process of language acquisition are all non-conscious processes. As far as language acquisition is concerned, this gets us back to where we started, since mentalese is hypothesized to exist at the subpersonal level anyway, we may very well have to accept there are some cognitive processes that cannot be explained by appealing to natural language. But this should not pose a problem for the thesis that I am defending because I am not arguing for the extreme claim that *all* cognitive processes should be explained in terms of natural language, as there are obviously some types of thinking that do not require linguistic representation. If anyone were to be guilty of arguing for such an extreme claim, it would be Fodor, supposing that he were to find explaining thinking-how in terms of mentalese appealing. Although this would suffice as a rebuttal, the result would extend the mentalese hypothesis beyond explaining linguistic phenomena and propositional attitudes to supposing that mentalese must underpin most, if not all, cognition in animals and human beings. The justification for this bold move is wanting, especially since the modest proposal that human beings represent their environment in natural language works just as well for all phenomena relating to propositional attitudes that require explanation, save language acquisition. And as we have seen, there are other models of language acquisition that do not necessitate the

postulation of an additional representational system; language acquisition as learning/thinking-how; perhaps a modernized and rehabilitated behaviourism; or more likely, connectionism. It is worth noting that connectionist systems have gained ground recently in explaining how language might be learned mechanistically according to set algorithms (Clark 2001: 63), without any representational processes requiring hypothesis formation and confirmation.

It may be that all of our thoughts are somehow encoded in mentalese. Since we are not conscious of the operation of mentalese, there is no introspective evidence to the contrary. But the burden of proof is on Fodor and other fans of the mentalese hypothesis to show that we have good reason to hypothesize the existence of mentalese. In this chapter I have argued that we have no good reasons to hypothesize the existence of mentalese, since natural language explains all the phenomena meant to motivate the postulation of mentalese, save language acquisition. We may still have good grounds to insist that conscious propositional thought is represented in the medium of natural language, but we may not be able to explain how natural language is acquired or other cognitive processes that occur outside the realm of consciousness without appealing to some sort of non-conscious mental activity, be it computations in mentalese or perhaps a connectionist model. We may be able to hitch natural language to a different explanatory horse than mentalese, but one thing is for sure, postulating natural language as a representational system will not explain how we acquire natural language. This shortcoming of the cognitivist position pales in comparison to the communicativist failure to discover anything resembling the

operation of mentalese in our brain activity, the implications of which I will now turn my attention.

4. The Epistemic Objection to the Mentalese Hypothesis

There are three arguments against the mentalese hypothesis made on the basis of the lack of empirical evidence for its existence. These arguments all make the same basic objection: there has been no success thus far in discovering the operation of mentalese in the mind/brain and given certain conditions obtain, there may never be any. There is controversy over whether this constitutes a valid criticism of the mentalese hypothesis. Fodor remarks that “ontology is one thing, epistemology is quite another” (Fodor 2003: 5), and draws our attention to the fact that the truth of a theory is quite different from whether we have the means of testing it. He is right, but Fodor neglects to mention other pragmatic virtues of scientific theories such as simplicity, fruitfulness, and falsifiability that the mentalese hypothesis lacks, especially when compared to the hypothesis that natural language is a medium in which we think.

In order to discover the existence of mentalese we must somehow abstract away from the brute neural activity to determine the functional role of such physical structures within the cognitive economy of mentalese. The neuroscientist in search of mentalese is trying to figure out how sentences in mentalese are causally connected to each other and to certain sensory input such that they produce certain behavioural output (Bermudez 2003: 28-29). The question is how a neuroscientist might go about

doing this. Bermudez is so skeptical that he questions whether it is even possible in principle. Below is a summary of the argument that he provides for his skepticism:

1. Either we identify physical structures instantiated by neural activity in terms of their semantic contents or syntactic features.
2. We cannot identify them in terms of their semantic contents because this assumes that we already know how to determine the content of a sentence in mentalese, which is what we are trying to explain.
3. The other half of the disjunction is not a viable option either because we cannot identify the syntactic features of the physical structures in question without at least having a basic knowledge of their semantic features. We cannot form hypotheses about the syntactic features of a language without making assumptions about what its words mean, since the syntax of a language provides instructions on how to fit words in different grammatical categories together. And it is indisputable that whether a word is in the grammatical category of verb, adverb, adjective or noun depends on what it means.

(C) We are at a loss as to how to tell relevant causal properties from the irrelevant causal properties in the chaos of brain activity that demarcate the syntax and the semantics of sentences in mentalese.

The benefit of supposing that natural language is the representational system that the neuroscientist should investigate is that she already knows the syntax and semantics of the language that she is looking for and the areas of the brain (the aural, vocal, and connected subsystems) in which to look (Hauser 1995). The neuroscientist

searching for mentalese on the other hand has no antecedent knowledge of the syntax and semantics of the language that she is trying to find. Hauser aptly compares the task of the former neuroscientist to the decryption project of Allies, and the latter to the decryption project of the Axis during WWII (Hauser 1995). The Allies were able to crack the German encryption code with the help of antecedent knowledge of German syntax and lexicon, whereas the American practice of translating messages to Navajo prior to encryption famously frustrated the Axis' code breaking efforts. This analogy is a stark illustration of which research program could potentially prove to be the most successful.

We may be wrong to dismiss the mentalese hypothesis simply because neuroscience has not yet figured out how to determine whether mentalese functions as a system of representation for our cognitive processes. But as I have shown in this chapter, the hypothesis that natural language is a system of representation has equal explanatory power, and as the example above illustrates, it also has greater potential for successful scientific confirmation. Furthermore, the idea that our cognitive processes occur in natural language has a beautiful simplicity in its explanation of thought as internalized speech, and it does not require the unparsimonious postulation of a representational system that is introspectively, and at the moment, scientifically, inaccessible. Therefore, when we consider the fact that accepting the theory that natural language plays a cognitive role as the medium of our conscious propositional representations reflects the scientific virtues of parsimony, fruitfulness, and explanatory power, we should reject the mentalese hypothesis in favour of the natural language as a language of thought hypothesis.

Chapter 2:

Natural Language as the Language of Conscious Propositional Thought

The purpose of this chapter is to offer an evaluation of the tenability of two arguments for the claim that the language in which we conduct some of our thinking is natural language. The first argues from introspection that we have good reason to believe conscious propositional thoughts are sentences of natural language presented to the mind through the auditory imagination. The second argues that there is no plausible medium other than natural language that is a suitable vehicle for reflexive thinking, or thinking about thoughts. These arguments share the presumption that there is something special about linguistic thinking. Language makes it possible to conceptualize and evaluate the content of consciousness, to think about an experience *as* being such-and-so. This self-consciousness is what sets linguistic thinking apart from nonlinguistic thinking and marks the difference between sapience and mere sentience. After presenting these two arguments, I will consider two objections and show how the introspective argument can successfully answer both of them.

1. Carruthers' Argument from Introspection

Carruthers' introspection argument derives much of its support from the intuitive plausibility of the notion that the majority of our conscious propositional thought occurs in the language that we use to communicate in our everyday lives. To make sure that everyone checks the plausibility of this by introspecting the same thoughts, some stipulations are in order.

First, we must introspect at the level of sentences and not individual words. If we introspect the meaning of the word 'cat,' we may form a mental image of a cat. However, remembering what Frege taught us, we will recognize that the meaning of a word can only properly be understood in terms of its contribution to the meaning of a sentence as a whole, and a word only has meaning within the context of an entire sentence. In addition, a largely imagistic account of thought (expounded by Locke, 1690, Hume, 1739, and Russell, 1921) founders on abstract concepts, the inferential structure of thinking and the meaning of the logical connectives that bind it together. Our mental images depict how we think something would look, but our thoughts contain many concepts that images fail to capture. Images do not capture the meaning of temporal concepts like 'yesterday' or 'next week' or abstract properties such as 'prime number' or 'vague' (Carruthers 1996: 32). It is unclear how mental images could successfully reflect the inferential structure of thought, especially since there are no mental images that express the meanings of negation, conjunction, disjunction, or the conditional. There are no images that adequately illustrate the meanings of 'and' or 'if...then' or explain how some thoughts are logically linked to others. We may think in images at times (for example, consider the artistic endeavor of the photographer, painter, or sculptor), but these thoughts are not propositional or linguistic. However, we may embed images within the context of propositions to form hybrid propositional-imagistic thoughts (Carruthers 1996: 36). Examples of this are found in thoughts such as 'We should move the couch over *there*' or 'If I rotate my suitcase like *this* it will fit in the trunk,' where an image takes the place of an indexical.

Second, the introspective argument applies only to conscious thoughts, which are naturally episodic in nature. For a thought to be conscious it must occur at a specific point in time. There are also propositional thoughts that are not present to consciousness, such as the belief that Ottawa is the capital of Canada or the desire that the Yankees not win the World Series. Carruthers refers to these thoughts as standing states. I may believe that Ottawa is the capital of Canada but not be conscious of that fact until some question or mental association brings it to my attention. We do not have transparent access to our standing states, as we often forget certain beliefs, desires or intentions. However, we may suppose that the introspective thesis covers standing states insofar as they potentially may become activated and thereby occur as conscious propositional thoughts.

Third, the introspective argument applies only to propositional thoughts. The negative reasons for understanding natural language sentences as being constitutive of thought were given above. To repeat, the imagistic account of thought is inadequate to the task of explaining the inferential complexity, abstract nature, and logical structure of our thought. Additionally, the thesis that thoughts are generally not imagistic but relations to internal sentences also receives positive support from the causal nature, the systematicity, and the productivity of thought.

To sum up, the target of Carruthers' introspection is conscious propositional thought, which is best understood in terms of sentences rather than images. The content of propositional thought follows the that-clause in a sentence. Preceding the that-clause we usually find some sort of attitude such as hoping, believing, desiring etc. We may be able to say that we see that-x or taste that-y, but Carruthers is not

arguing that we perceive the world in natural language. Rather, he is suggesting that judgments about the nature of our experience (believing-that, supposing-that, wondering whether etc.) as being such-and-so are linguistic, such that when we think to ourselves 'Is that a bear behind that bush or just a lump of dirt?' we are deploying sentences of natural language.

Carruthers says that we should introspect two different types of conscious propositional thought, first, when we use a sentence publicly to communicate and second, when we entertain a thought or sequence of thoughts privately. In the first case, introspection in most cases will show that there is no separable mental process accompanying the expression of a sentence, at least not available to consciousness (Carruthers 1996: 50). In the second case, Carruthers says that introspection shows that private thoughts primarily take the form of sentences of natural language heard in the auditory imagination, which he sometimes calls 'imaged sentences' or 'phonological representations.' To put his thesis in a clear little slogan, "Inner thinking is mostly done in inner speech" (Carruthers 1996: 50).

According to Carruthers, the upshot of this argument is that if the medium of conscious thought is natural language, then the medium of conscious thought cannot be mentalese (1996: 55) at least if we rely solely on what introspection tells us. Carruthers concludes that this is evidence that Fodor's claim that all thinking occurs in mentalese is false. Introspecting our consciousness provides no support for the contention that we think in mentalese and then translate our thoughts into natural language so that we may communicate them. But criticizing Fodor on this ground is misleading. Fodor nowhere makes the claim that the mentalese hypothesis is about

thought at the personal level, the level of conscious awareness. Rather, it is supposed to be an explanation of what occurs at the personal level as a hypothesis about the *subpersonal* cognitive architecture that grounds it. Fodor does not suggest that we have conscious awareness of the operation of mentalese, nor does he suppose that a lack of introspective access to mentalese poses any sort of argument against its hypothesis. Bermudez's argument that some of our thinking occurs in the medium of natural language is stronger because it does not miss this crucial point. But that will be the subject of the next section. For now, I will leave this digression and return to the argument under consideration.

Carruthers' argument would be very weak if he relied only upon the introspection of his own conscious propositional thoughts, since some people might come to different conclusions by way of the same method. This is where he falls back on empirical evidence. Carruthers cites both colloquial evidence and scientific empirical data to support his claim that inner thinking is mostly done in inner speech.

First I will canvass the colloquial evidence. As I write this, my consciousness consists of an array of various formulations of different sentences that I could put down on the page. Introspection tells me there is no difference between sentences that remain thoughts and sentences that I choose to write down. This evidence could be considered a cheat, since writing is essentially a linguistic activity, but I could just as easily refer to what the content of my consciousness appears to be when I reason with someone else, or with myself about a course of action. In both cases, I rely on representations in the auditory imagination (what Carruthers and I call 'imaged sentences') of various possibilities to determine what I should do or say. For example,

I may say to myself 'Either I write my paper, or go to judo practice,' 'I really want to go to judo,' 'but this paper is due soon and judo will make me too tired to work on it tonight,' and so forth. If I cannot come to a conclusion, then I may vocalize this silent chain of reasoning to someone else in the hope that she can help me determine the best course of action. It is no accident that we sometimes describe our public monologues as 'thinking out loud.' For further evidence of this, consider the fact that people who learn a second language often report a key point in their education when they are able to think (and sometimes dream) in the language that they are studying.

The common thread in all this colloquial introspective evidence is that the only difference between public and private thinking is that in the former case a thought is uttered aloud and in the latter it remains an imaged sentence. When I reason about a problem with myself, or out loud with someone else, introspection shows that both activities are essentially the same in that they involve sequences of imaged sentences and their manipulation. Moreover, while writing at a computer, or on a blackboard, or having a conversation, there is nothing in my consciousness that indicates there is any difference between a thought and its expression. It feels as if the thought of a sentence is only different from its expression insofar as it is first heard privately as an imaged sentence before it is made public in writing or speech. However, the tenability of this entire argument rests on how trustworthy introspection is in informing us that the medium in which we think our conscious thoughts is the medium in which we speak. Both objections discussed in the third section of this chapter target this claim, which is crucial to Carruthers and Bermudez's arguments

that we conduct at least some of our thinking in the language that we speak. For now, I will set this question aside.

For scientific support of his introspective claims, Carruthers relies on the work of Russell Hurlburt (1990 and 1993), who has developed a method for studying inner experience, which he calls Descriptive Experience Sampling (DES). DES uses a beeper to cue subjects to record their experience at random intervals, which they then describe in greater detail during a follow-up interview twenty-four hours later. Carruthers finds the results of DES studies promising. All subjects describe at least some of their inner thinking as inner speech, although the frequency varies widely, from as low as 7% for some subjects to as high as 80% for others. Some subjects have reported wordless and imageless purely propositional thoughts, which creates a problem for Carruthers' claim that the vehicle of conscious propositional thought is natural language.

The short explanation of how Carruthers finesses this problem is that he argues that such thoughts are only *apparently* conscious, but really what is going on is a matter of "swift retrospective self-interpretation, much as if we are ascribing a thought to a third party" (1996: 56). The distinction between conscious and non-conscious thinking is an invaluable asset to Carruthers, both in criticizing Fodor's arguments for the view that all of our thinking is conducted in mentalese and for defending his own position against potential objections. I will discuss how Carruthers uses this distinction at length to respond to Nisbett and Wilson's argument against the reliability of introspection in the third section. Presently, it would be helpful to briefly explain Carruthers' understanding of consciousness. He says that unlike the thoughts

of a third-party, we are non-inferentially aware of our conscious thoughts. Moreover, the distinctive feature of conscious thought is that it is available to indefinitely reflexive thinking. Carruthers suggests that the fact that we can think about the fact that we are having a given experience is what makes it conscious and he ties this capacity for conscious thought to being a speaker of natural language. He is not alone in this supposition; Bermudez makes a short and sweet argument for the thesis that reflexive thinking is connected to natural language, which I will turn to next.

2. Bermudez's Argument from Elimination

Bermudez favours Carruthers' account of natural language as the medium of conscious propositional thought, but he takes a different route to this conclusion. Later, I will argue for the superiority of Bermudez's argument, because it evades objections that block Carruthers' line of reasoning. As his starting point, Bermudez uses Clark's supposition (1996) that the difference language makes to human cognition is that thought formulated in words only or in writing is altered insofar as it becomes an object both for ourselves and for others. As this sort of object, it differs from other types of thought because thought given linguistic expression becomes something stable that we can think further thoughts about. This makes reflexive thinking possible, which Carruthers says makes thought conscious. In contrast, Bermudez supposes that some reflexive thought is conscious, but he does not go so far as saying that reflexivity is what makes a thought conscious. I will remain agnostic about whether reflexivity *makes* thinking conscious, but I will defend the

view that reflexive thinking is, by (Clark's) definition, conscious thinking and that encoding propositions in natural language makes reflexive thinking possible.

Clark provides a helpful description of reflexive thinking, which he prefers to call 'second-order cognitive dynamics.' For the purposes of clarity, I will continue to use the term 'reflexive thinking' to describe what Clark elucidates so well in this passage as

a cluster of powerful capacities involving self-evaluation, self-criticism and finely honed remedial responses. Examples would include: recognizing a flaw in our own plan or argument, and dedicating further cognitive efforts to fixing it; reflecting on the unreliability of our own initial judgments in certain types of situation and proceeding with special caution as a result; coming to see why we reached a particular conclusion by appreciating the logical transitions in our own thought; thinking about the conditions under which we think best and trying to bring them about. The list could be continued, but the pattern should be clear. In all these cases we are effectively thinking about our own cognitive profiles or about specific thoughts (Clark 1996: 177).

Bermudez's slogan is "no intentional ascent without semantic ascent" (2003: 164) which is to affirm Clark's proposition that language enables us to think about our own thoughts. It is not exactly clear how fundamental reflexive thinking is to mentalese, because reflexive thinking is paradigmatically a conscious activity and the mentalese hypothesis is not an attempt to explain conscious thought. Bermudez supposes that supporters of the mentalese hypothesis would flatly deny this supposition, because they claim the proposition that reflexive thinking is available in mentalese is a fundamental aspect of the mentalese hypothesis (2003: 158). This is clear if we consider what is going on in the practices of infant language acquisition, in which infants are supposed to acquire natural language through a process of hypothesis formation and confirmation about the meaning of words.

Based on the definition above, the process of hypothesis formation and confirmation in infant language acquisition should count as reflexive thinking, but it

is not apparent whether this thinking is conscious. It would be most charitable to the mentalese hypothesis to suppose that language acquisition is a subpersonal process, because the supposition that infants are little field linguists actively, intentionally, and consciously constructing theories of interpretation is difficult to believe in the absence of strong empirical evidence to support it. This leaves us with a dilemma. We may suppose that the mentalese hypothesis' account of language acquisition is an example of reflexive thinking that is not conscious, or that it is not an example of reflexive thinking. As Clark describes it, reflexive thinking appears to be co-extensive with self-conscious awareness of our own mental states. Hence, I believe that Bermudez mischaracterizes the mentalese hypothesis by supposing that reflexive thinking is of any importance to it. For it is a mistake to say that reflexive thinking can occur at the subpersonal level as he does when he states "certain types of hypothesis testing and refinement do take place at the subpersonal level. Something like this happens, according to Fodor, when we learn a language" (Bermudez 2003: 159).

The bulk of the writing in Carruthers' *Language, thought and consciousness* is premised on the incompatibility of the idea that we think in natural language with the idea that we think in mentalese. Under the proviso that introspection is a reliable source of information about the nature of conscious thought, Carruthers argues that if our conscious propositional thinking occurs in our mother tongue, as introspection tells us that it does, "*then it cannot occur in any other natural language, let alone in a supposedly innate, universal, symbolic system such as mentalese*" (Carruthers 1996: 55, emphasis added). Carruthers does not have license to conclude that introspection

gives us some reason to think that conscious propositional thoughts occur in natural language and not mentalese, since mentalese is not postulated on the basis of introspection but rather on the basis of its power to explain subpersonal phenomena such as language acquisition and other thought processes in prelinguistic and nonlinguistic organisms. To repeat, I believe that it does no harm to the mentalese hypothesis to suppose that it operates entirely at the subpersonal level. Thus, the opposition between the proposition that we think in natural language and the proposition that we think in mentalese is imaginary. There is no reason to think that these propositions are mutually exclusive if we suppose that we think in natural language at the personal level and mentalese at the subpersonal level.

Bermudez is also guilty of attempting to create a false debate about whether mentalese or natural language serves as the vehicle for conscious cognitive processes. He says that it is hard to conceive a defender of mentalese supposing that an infant's reflexive thinking in mentalese is conscious, while maintaining that the operation of mentalese in general is part of our subpersonal cognitive architecture (2003: 159). I doubt that there is a defender of mentalese that holds such an inconsistent position, and Bermudez appears to be stretching to invent one for the sake of argument. He acknowledges as much when he goes on to say that the generally accepted prevailing view is that "subpersonal states are inferentially insulated from conscious processes of cognitive evaluation and self-criticism" (2003: 159) like those described in Clark's account of reflexive thinking. We may conclude that since the mentalese hypothesis is a hypothesis about subpersonal cognitive architecture, and conscious access to our

thoughts is an essential aspect of reflexive thinking, then it follows that reflexive thinking cannot occur in mentalese.

It is worthy of note that there are some people who have suggested that the mentalese hypothesis need not be a hypothesis solely about subpersonal cognitive architecture. It has been proposed (Rosenthal 1991) that there is some kind of internal language that thoughts at the subpersonal level are translated into before they reach the level of conscious consideration. This suggestion would fit well with advocates of higher-order thought theories. Bermudez says that we should reject this suggestion on the basis that our phenomenology does not offer it support. He agrees with Carruthers that introspection tells us that conscious propositional thoughts occur in sentences of natural language. We may have feelings and visual images that are not sentential, but he insists that the evidence from introspection should be trusted; he goes so far as to say that we never have conscious propositional thoughts without linguistic vehicles. Furthermore, natural language provides the only plausible medium for thought that is to be the target for reflexive thinking. This claim is defended in the reconstruction of his argument in the next paragraph and its soundness is dependent on the exhaustive consideration of alternative forms of representation to natural language. I believe that Bermudez thoroughly considers the alternatives and provides good reasons to reject them and so we should accept the soundness of his argument.

The following is a representation of Bermudez's argument that reflexive thought is linguistic. Thus far, we have covered steps (1)-(4). What needs to be demonstrated next is that language is the only plausible target for reflexive thinking at the personal level.

1. The only way that reflexive thinking is possible is if thoughts have vehicles that allow them to be the objects of further thoughts.
2. The vehicle for reflexive thought must either be personal or subpersonal.
3. Subpersonal vehicles, such as mentalese, do not offer the conscious access that is a necessary feature of thinking about our thoughts.
4. Therefore, reflexive thought must occur at the personal level.
5. There are three types of vehicles available at the personal level for reflexive thought. They are mental maps, mental models, and natural language.
6. Mental maps do not offer the inferential flexibility or logical structure that characterizes reflexive thought.
7. Mental models do not provide a genuine alternative to natural language as a vehicle of reflexive thought, since they only provide an imitation of the propositional structure of reflexive thought.

(C) Therefore, reflexive thought must occur in the medium of natural language, because natural language provides the only plausible vehicle for its occurrence.

Bermudez's objections to the supposition that mental maps (proposed by Baddon-Mitchell and Jackson 1996) are vehicles for reflexive thought should be familiar, as mental maps refer to a variety of mental image. Hence, Bermudez's objections to this view are similar to Carruthers' difficulties with the proposal that conscious propositional thought is imagistic. Mental maps are structurally isomorphic pictorial representations of states of affairs. Mental maps, much like actual maps, capture the spatial relations between the objects in the world that they represent. A

mental map of *the cat is on the mat* would represent the physical relationship between the cat and the mat. Considered in this context, mental maps seem like plausible vehicles for reflexive thought. However, reflexive thinking requires that its vehicles represent the *inferential relations* between thoughts, and there has yet to be a convincing explanation of how images could possibly represent inferential connections (Bermudez 2003: 161). There is potential that certain transitions in reflexive thinking could be represented in terms of associations between images, such as reconsidering a belief and determining that it was accepted without sufficient evidence. Even if these basic transitions do not pose a problem for mental maps, they still remain inadequate for representing more complex transitions in reflexive thinking, like transitions dependent on an understanding of deductive validity or probabilistic support. There is no pictorial equivalent of these concepts, or even of basic logical concepts such as negation or the conditional. Bermudez suggests that for maps to work for reflexive thinking we have to resort to propositional terms, understanding maps as propositions and then evaluating their inferential connections. And in doing so, we adopt the linguistic model of reflexive thinking that Bermudez is advocating.

Bermudez argues that in order to explain our practices of reasoning mental models must also adopt a linguistic model of reflexive thinking. The motivation for mental models (proposed by Johnson-Laird and Byrne 1991) is that they supposedly offer an explanation that is truer to the empirical information on how people reason. Instead of relying on sentential representations as the objects of reflexive thinking, mental models of arguments are built out of relevant premises and validity is cashed

out in terms of whether an argument holds in all models, probabilistic validity in terms of whether it holds in some models, etc. Bermudez's contention is that although mental models are structurally isomorphic to the states of affairs, the states of affairs that they represent are derivative of the premises of the arguments they are modeling. These premises are linguistic entities. Mental models depend on understanding inference as transitions between sententially encoded propositions that are represented as structurally isomorphic models of the states of affairs expressed by the propositions under consideration. Hence, mental models only provide a way of understanding sentential reasoning, not a genuine alternative to it. Therefore, with all its competitors eliminated, the only plausible medium for reflexive thought is offered by the sentences of natural language. These sentences allow for thinking about thoughts because they make it possible to think about the sentences through which our thoughts are expressed (Bermudez 2003: 164).

3. Objections

Carruthers and Bermudez both rely on introspection to infer the linguistic nature of our conscious propositional thoughts. However, there are important yet subtle differences in their respective arguments. Bermudez does not rest his argument solely on the strength of introspection. Instead, he makes the claim that reflexive thinking requires the medium of natural language, as no other medium could make this type of thinking possible. Reflexive thinking is an example of conscious propositional thought, but it does not follow that all conscious propositional thought is reflexive. Conscious propositional thought is potentially reflexive, in that it may be subject to

reflexive thinking; once encoded in natural language it becomes the object of further thoughts. In this way Bermudez's argument makes a slightly weaker claim. Bermudez supposes that all conscious propositional thought is potentially reflexive and as such natural language is potentially constitutive of all conscious propositional thought, but he does not go as far as Carruthers who argues that all conscious propositional thought is represented in the medium of natural language.

Not surprisingly, Carruthers' argument is on shakier ground. It rests on the dubious inference from the proposition that

1. Introspection tells us that the content of our conscious propositional thought has the property of being linguistic.

to

2. The vehicle of our conscious propositional thought must have the property of being linguistic.

This inference depends on the assumption that what introspection tells us about the properties of the content of our thought is also applicable to its vehicles. Machery (2005) argues convincingly that this inference is fallacious and on this basis introspection cannot constitute evidence for the linguistic nature of thought. As we shall see, Bermudez is able to avoid this objection because he argues that natural language functions as the vehicle of reflexive thought without making the inference that the thought it is about — the content of the thought — is linguistic.⁶

Before considering this objection any further I will investigate the soundness of inferences from introspection of our phenomenal experience to conclusions about the nature of thought. This is extremely important for Carruthers because the

⁶ I owe this point to Dr. Patrick Rysiew.

reliability of introspection forms the lynchpin of his argument. It is important for Bermudez as well, but if introspection does not show that all conscious propositional thought is represented in the medium of natural language his argument may still go through, as he is arguing that only reflexive thinking involves sentences of natural language. Carruthers and Bermudez reject the idea that any of our conscious propositional thoughts occur in mentalese prior to our conscious consideration as something that lacks “phenomenological plausibility” because we are not consciously aware of thoughts being translated from mentalese into natural language. Our phenomenology simply tells us that our conscious propositional thought occurs in the medium of natural language. The question is whether introspection of our phenomenal experience is a reliable source of information about the nature of our cognitive processes.

Nisbett and Wilson

Christopher Gauker answers the above question dismissively, saying that ‘we all know it has been demonstrated that introspection is unreliable’ (1994) and cites Nisbett and Wilson’s 1977 as proof. This conclusion is far too strong, and although Nisbett and Wilson provide some empirical data to show that introspection does not always provide us with transparent access to our mental processes, this does not warrant dismissing introspection as a source of information about them altogether. Furthermore, Carruthers actually uses Nisbett and Wilson’s findings to bolster his introspective argument.

There are a variety of different studies quoted in Nisbett and Wilson’s 1977, but I will only discuss the ones that are most relevant to Carruthers’ case. In these

studies, subjects were asked to describe their thought processes in making various decisions, but their reasons proved to be out of touch with the reasons for their behaviour that were obvious to any outside observer. First, subjects were asked to choose a favourite from an array of identical items, and people showed a much stronger preference for items on the right side of the array. However, when asked for a reason for their choice, subjects would not cite this right-side preference, but would instead refer to things like superior quality, appearance, etc. Secondly, a study found that people who were paid to play with a puzzle reported less intrinsic interest in the activity than those who played with it voluntarily, but these reports failed to match up with the amount that they played with the puzzle in their free time. Third, people are often poor judges of the factors that influence their evaluations or decisions, such as which aspects of a job-applicant's portfolio influenced their decision to call him or her for an interview. Moreover, people told of such studies make the same poor judgments of their real motivations to call someone as the subjects in them.

Carruthers concludes that the best explanation of these data (and this is a weaker version of the conclusion drawn by Nisbett and Wilson) is that subjects in these circumstances lack any conscious access to their thought processes. Instead, Carruthers posits that the subjects rely upon swift retrospective self-interpretation that allows them to attribute to themselves thoughts that they believe that they should have such that their behaviour and responses could be understood as reasonable.

How does this strengthen Carruthers' introspective argument? If you recall, Hurlburt reported a significant number of occurrences of purely propositional (non-verbal) thoughts available to consciousness in his DES studies. This potentially

undermines Carruthers' claim that introspection gives us good reason to believe that all our conscious propositional thoughts are linguistic. Carruthers uses his explanation of what is going on in Nisbett and Wilson's empirical studies of introspection to finesse this information. He suggests that when people suppose they have had a nonlinguistic conscious propositional thought, that thought is actually the result of "swift self-interpretation" (Carruthers 1996: 241). And thoughts that require self-interpretation are not conscious thoughts, because we are non-inferentially aware of our conscious thoughts. Interpretation takes place only when we do not have non-inferential access, like when we ascribe propositional attitudes in order to explain the behaviour of others. Although these retrospective self-interpretations may feel like the other thoughts to which we have non-inferential access, Carruthers counters that we often do not feel we are interpreting the behaviour of others either, we just see that they are in a certain mental state, such as anger or despair, even in the context of theatre when we actually know we are observing characters acting.

Carruthers successfully saves his position from the objections to introspection offered by Nisbett and Wilson, albeit with *ad hoc* stipulation that appears to be more sensitive to his own position than the introspective reports of the test subjects canvassed by Nisbett and Wilson or the DES studies. In any case, Carruthers puts his reply to the objections of Nisbett and Wilson to good use in defending his introspective argument against the contradictory data of Hurlburt's 1993. I do not take Carruthers' dodge as being definitive, but he does offer sufficient conditions for his position to be plausible, if not necessary. However, Nisbett and Wilson's

criticisms of introspection are not nearly as difficult to counter as Machery's misgivings.

Machery

According to Machery, Carruthers commits something he calls 'the vehicle/content fallacy' (2005: 483). Machery describes this fallacy as inferring that the vehicle of thought must possess the property p , given that the content of thought possesses the property p . An example of this is inferring that because the content of my mental image of the sky is blue, then the vehicle of that thought should also be blue. Carruthers appears to be guilty of doing exactly this, as I described in the inference his argument from introspection makes in moving from (1) to (2) above. However, this does not necessarily give us good reason to believe Machery's claim that introspection cannot proffer evidence about the role of natural language in thought. This will be borne out by Bermudez's argument with which I will counter Machery's charge that "the introspective fact of inner speech cannot be evidence that our thoughts are linguistic" (2005: 473). He calls this the Blindness of Introspection Thesis (BI) and he offers the following reasons why we should accept it:

1. Carruthers' conclusion that conscious propositional thought occurs in the medium of natural language concerns the vehicle of our thought.
2. We do not have any introspective access to the properties of the vehicles of our thoughts.
3. Hence, the introspective fact of inner speech cannot be evidence that our conscious thoughts are linguistic (BI) (2005: 474).

Obviously the weight of the argument rests on the truth of (2). Machery says that we have reason to accept (2) on the basis of mistaken conclusions about the vehicles of our imagistic thoughts on the basis of their content. The idea that the content of our mental images can tell us anything about their vehicles has fallen out of favour. "For

example, it is widely recognized that the visual image of a red apple does not have to be red: true, the visual image represents the property red, but that fact does not support the claim that the vehicle of this image also has that property” (476).⁷

Carruthers provides strong arguments to the effect that sentences of natural language form the *content* of our conscious thoughts, so it is unnecessary for me to restate the evidence of this that Machery provides. Nor is it necessary for me to restate Carruthers’ conclusion that the vehicles of our conscious propositional thoughts are sentences of natural language, which he reaches on the basis of the fact that introspection of the content of our consciousness tells us that it is so. The question is whether we should reject Carruthers’ argument on the basis that it duplicates the mistakes of philosophers such as Locke and Hume, who infer that the vehicles of our thought must be mental images on the grounds that the content of our thought appears to be mental images.

Before we make too hasty a judgment about the answer to this question, we ought to give consideration to the *explanandum* of the causal efficacy of our propositional thoughts. We are introspectively aware of the power of auditory representations of natural language sentences to cause us to infer the truth of other sentences. For example, I may think ‘If it is raining, then the streets are wet,’ and upon noticing that it is raining outside, conclude that ‘The streets are wet.’ Valid patterns of reasoning such as this occur in our thought processes all the time. This phenomenon does not concern the linguistic content of our thoughts individually, but rather the syntactic properties of their linguistic vehicles that allows for logical

⁷ Wittgenstein makes a similar observation about the meaning of ‘cube’ (1953: R. 74 & 141). The image is inessential; it is the correct use the word connected to the image that is important.

relationships to obtain between them. The fact that we can introspectively observe that our linguistic thoughts can be causes of both actions and other linguistic thoughts requires an *explanans*, the most obvious being that we have introspective access to at least some of the properties of the vehicles of our thoughts.

Fodor and Machery cite Humean reasons to halt the search for an *explanans* within the realm of introspection, I believe that should be sufficient to block any reply meant to salvage Carruthers' introspective argument. It does not follow from the constant conjunction of P and Q that P causes Q. The fact that sentences of inner speech respect valid logical transitions does not prove that the sentences are causally responsible for this fact. As Fodor eloquently puts it, "causal connectedness can't be perceived, it has to be argued for" (1998: 64).

Bermudez persuasively argues for the causal connection between the inferential transitions among the sentences that constitute our reflexive thinking. As stated above, Bermudez's argument relies on the claim that reflexive thinking is made possible because its vehicle is linguistic, while making no claim that the content of reflexive thought must be. Reflexive thinking is thinking about thought. The vehicle for reflexive thinking does not need to share the property of the content that the thinking that it is about. So this allows Bermudez's argument to go through. Furthermore, examination of other conceivable vehicles for reflexive thinking above has shown that natural language is the only plausible vehicle for reflexive thinking. Thus, Bermudez's argument does not commit the vehicle/content fallacy that foils Carruthers' argument.

In this chapter I have discussed Carruthers and Bermudez's arguments that at least some form of thought must occur in natural language. Consideration of the objections to Carruthers' argument has led me to the conclusion that those wishing to defend the role of natural language in our cognition ought not to rely solely upon introspection to decide the issue, as the success of Bermudez's position in withstanding the same objections demonstrates.

Chapter 3:

The Flaws with Fodor's Understanding of Meaning

To recap and rephrase some of what was said in the introduction, the mentalese hypothesis is meant to explain our cognitive processes in such a way as to provide a scientific vindication of common sense folk psychology. We use folk psychology to explain the behaviour of our fellow humans and the intelligent behaviour of some animals. This involves the ascription of propositional attitudes, such as beliefs, desires, fears, hopes and so on. For instance, if we can see that someone desires Q, and we know that she knows that if P, then Q, we can predict, *ceteris paribus*, that she will try to bring it about that-P. The mentalese hypothesis supposes that when we ascribe a propositional attitude, we are doing more than just interpreting an agent's behaviour — we are actually describing her attitude toward the tokening of a representation in mentalese. The natural language as a language of thought hypothesis supposes instead that when we ascribe a propositional attitude, we are actually describing an attitude toward the tokening of a natural language sentence.

We use natural public language to ascribe propositional attitudes, and in doing so we describe the mind of the person whose behaviour our propositional attitude ascriptions are supposed to explain. The theory that natural language sentences constitute the representations described by folk psychology is a more accurate description of what the 'folk' take themselves to be doing, since it is likely they have never heard of Fodor's mentalese. Whatever the case, both the mentalese hypothesis and the natural language as a language of thought hypothesis are committed to the

scientific vindication of folk psychology; they suppose that we will one day discover the physical entities in the mind/brain we refer to with our everyday language of propositional attitude ascriptions. This is where philosophy of mind and philosophy of language are supposed to join hands and offer a unified explanation of meaning in mind and meaning in language. This has not happened, however. In fact, there have been arguments made about meaning in the philosophy of language that jeopardize the viability of a representational theory of mind. For this reason, Fodor wrote *Psychosemantics*, in which he takes on arguments that make “inferences from premises in the philosophy of language to skeptical conclusions in the philosophy of mind” (1987: xii). Fodor argues that the majority of the theories of meaning accepted in the philosophy of language, by accepting some form of holism, deny the possibility of intentional realism. In order to be an intentional realist, one must take thoughts to be discrete, compositional, semantically-evaluable, and causally-efficacious states. In other words, distinct thoughts have distinct physical realizations. Thoughts are composites of parts that can be shared with other thoughts; they can be true or false; and they can cause other thoughts or behaviour.

The purpose of this chapter is to argue that contrary to what Fodor says, we can be intentional realists without being semantic atomists. Fodor thinks that the only way for a theory of meaning to be compatible with intentional realism is for it to be atomistic. And causal co-variance semantics offers the only plausible atomistic approach to meaning, which is only plausible as a theory of content for the mentalese, not natural language. So as far as Fodor is concerned, if you are an intentional realist, then you have to buy his theory of content, which means you are also buying his

mentalese hypothesis. In this chapter I want to show that we can accept the truth of intentional realism without buying anything that Fodor is selling. I will do this in four steps. First, I will give an exegesis of Fodor's arguments against theories of meaning derived from analyses of natural language, which he says subscribe to a holistic conception of meaning that he supposes to be incompatible with intentional realism. Second, I will discuss Fodor's atomistic alternative. Third, I will show that Fodor's alternative results in at least as many problems as it claims to fix. And fourth, I will offer an account of holism that preserves intentional realism.

1. Fodor's Attack on Holism and Functional-role Semantics

Fodor thinks that meaning holism gives the best argument against intentional realism. If meaning holism is true, Fodor says that propositional attitude psychology has no hope (1987: 56). This is why. According to Fodor, meaning holism individuates an intentional content in terms of the connections between a proposition and the *totality* of other propositions that the agent in question believes are epistemically relevant to its meaning. Fodor calls the propositions relevant to a proposition's content its epistemic liaisons. He stresses that epistemic liaison is a psychological notion, not an epistemological one. The meaning of a proposition is not objective; it's dependent on what liaisons an agent thinks are relevant to understanding what a given proposition means.

For example, consider my belief in the proposition that *the Chicago Bears will win the Super Bowl*. In order to understand this proposition, we must have a conception of what it would be like for it to be true. From this it follows that there are

also many other propositions whose truth or falsity must also be relevant to understanding the proposition that *the Chicago Bears will win the Super Bowl*. Propositions such as *the best team always wins the Super Bowl* and *the Bears are the best football team*. If someone thought that *the Chicago Bears will win the Super Bowl*, but she did not have these particular epistemic liaisons, (suppose that she thought that the Bears are a hockey team) then we would deny that her belief that *the Chicago Bears will win the Super Bowl* has the same intentional content.

The problem emerges when we try to determine the criteria by which to identify propositional attitudes with the same intentional content. It is obvious that people will vary widely in their opinion of what is epistemically relevant to understanding the content of a proposition. Not only that but meaning holism holds that it is the *totality* of epistemic liaisons that are relevant to the individuation of an intentional state. Consequently, we would be extremely lucky to find two people, even just two time slices of the same person, which we could say are in the same intentional state. The end result is that we cannot make the intentional generalizations that are necessary for a scientific treatment of propositional attitudes. For this reason, Fodor says that we must choose between intentional realism and meaning holism. He thinks that since meaning holism is in no way common sense (unlike the tenets of intentional realism that emerge out of common sense psychology), all he has to do is show that there are no solid arguments showing that it is better to be a meaning holist than an intentional realist. In the case of a draw, intentional realism should get the benefit of the doubt because it pleases our most common sense intuitions.

The dilemma that Fodor is facing us with is false, and in the closing section of this chapter I will describe how a holistic account of meaning can in fact be compatible with intentional realism. The next step in addressing Fodor's argument that holism is incompatible with intentional realism is to treat it in more specific detail, as it applies to functional-role semantics. The argument offered so far is only a sketch of the general form that Fodor says is taken by all arguments for meaning holism. If you missed it, Fodor says this ur-argument for meaning holism proceeds in three simple steps:

Step 1: Argue that at least some of the epistemic liaisons of a belief determine its intentional content.

Step 2: Run a 'slippery slope' argument to show that there is no principled way of deciding which of the epistemic liaisons of a belief determine its intentional content. So either none does or they all do.

Step 3: Conclude that they all do (1,2: *modus tollens*) (1987: 60).

Fodor argues that once we accept functional-role semantics, or any semantics that accepts that some epistemic liaisons of a belief determine its intentional content, then we take a step that inevitably leads us down a slippery slope to intentional anti-realism. Fodor believes that the problem with meaning holism is that it makes it impossible to individuate propositional attitudes. The upshot is that we can then no longer understand intentional states as discrete from each other, so we cannot distinguish an individual thought's semantic content or physical realization. This crushes any hope for scientific psychology and both the mentalese hypothesis and the natural language as a language of thought hypothesis. The question is why this theory of meaning became so popular when it is potentially so damaging to the intentional realism that many take to be common sense.

According to Fodor, the best explanation for why functional-role semantics came to prominence is because it admirably addresses the difficulties faced by simple denotational theories of meaning. Fodor actually thinks that the denotational theory of meaning is “*very nearly true*” (1987: 73), but only when understood as a theory of content for mentalese. The main problem with denotational theories is that they slice meanings/concepts/mental states too thick. The result is that in cases like that of Oedipus, the denotational account of meaning is inadequate for the provision of a psychological explanation. Since Oedipus wanted to marry Jocasta and not Mother, but since both names denote the same person, then as far as the denotational theory of meaning is concerned the propositions ‘I want to marry Mother’ and ‘I want to marry Jocasta’ are synonymous for Oedipus because they refer to the same object. But if we were to ask Oedipus about this he would deny the first proposition and assent to the second, so there must be something wrong with how the denotational theory individuates meanings because it is not sufficient for the provision of satisfactory psychological explanation.

This is where functional-role semantics steps in and explains that Oedipus’ Mother-thoughts are in fact distinct from Oedipus’ Jocasta-thoughts. Although both have the same truth-conditions, *vis-à-vis* being satisfied by the same object, functional-role semantics can separate their meanings by virtue of the fact that they play different roles. Functional-role semantics delineates different functional roles by different epistemic liaisons, and by doing this it may successfully account for the different identity statements Oedipus knew to be true of Jocasta. Hence, for Oedipus

'incest is improper' would be an epistemic liaison to thoughts about marrying Mother and not to thoughts about marrying Jocasta.

Although this may sound good, Fodor warns that it is too good to be true. This is because something very important is lacking from the explanation in the preceding paragraph: how do we cash out content in terms of functional role? The intentional vocabulary used for individuating Oedipus' various intentional states in the preceding paragraph is not available for the individuation of functional roles. For functional-role semantics to avoid begging the question, functional roles must be picked out non-semantically and non-intentionally (Fodor 1987: 76).

To solve this problem, Fodor thinks that functional-role semantics draws a moral from the network of causal interrelations among mental states to the supposition that the semantic interrelations among propositions produce a network of their own. These two networks are made up of the logical and semantic interrelationships between propositions or intentional states and their respective entailments. For example, the proposition that it is raining likely entails that the streets are wet, that it is overcast, and that it is not sunny. These entailment relations are partially constitutive of the meaning of the proposition that it is raining. This proposition and its entailment relations are isomorphic with potential intentional states such as believing that the streets are wet, believing that it is overcast, and believing that it is not sunny.

This is how I think Fodor suggests functional-role semantics goes from here: By assuming that there is isomorphism between these networks, we can individuate the semantic content of propositions like so. First, we limit our ascription of

propositional contents to mental states that display a pattern of causal relations such that we maintain the isomorphism between the causal and semantic networks. Second, we carry this out by making sure that the propositional contents we ascribe display a pattern of causal relations appropriate to their propositional contents. The idea is that we individuate the semantic content of propositions in virtue of the causal role of an agent's attitude toward some extensional object and assess the probability of an agent being in a certain intentional state given her attitude.

Fodor explains that the challenge for this approach to functional-role semantics emerges when we consider Putnam's Twin Earth thought experiment (Putnam 1975), which throws a monkey wrench into the explanation of the relationship between a mental state individuated by its functional role and its denotation. This is how. Suppose that there is a Twin Earth identical in every way to this one, except that on Twin Earth the stuff people call 'water' has the chemical composition of XYZ, not H₂O. This means that although 'twater' plays the same functional role in my twin's mental economy, when he says 'twater' he does not mean H₂O, he means XYZ. The upshot is that the content of an intentional state cannot determine its denotation.

As a solution to this problem defenders of functional-role semantics have introduced two-factor theories (Block 1986). These theories treat content and denotation as independent of one another. As explained above, the isomorphic connections between the causal network of mental states and the semantic network of propositions determine content (factor-1), whereas denotation is determined by the causal connections between the world and our thoughts (factor-2). This is how

functional-role semantics proposes to distinguish thoughts about water and thoughts about twater in terms of their respective causal connections to H₂O and XYZ. But the propositional content of mental states individuated in factor-1 must have satisfaction conditions, which depends essentially on the structure of the world we are thinking about. So then how are we to distinguish factor-1 from factor-2? If my dopplegänger is thinking about twater and I am thinking about water, there appears to be no way to tell whether the propositional contents of our thoughts are the same or different. For our thoughts possess the same functional role but have different causal connections. My thought that 'water is wet' is true if and only if water (that is, H₂O) is wet, whereas my dopplegänger's thought that 'twater is wet' is true if and only if twater (that is, XYZ) is wet. On the one hand, when we ascribe content in terms of functional role, we have the same thought. On the other hand, when we ascribe content by connecting these thoughts to their denotation, we have different thoughts. For this reason Fodor concludes that the functional role approach to content is doomed. Fodor can see no way for functional-role semantics to ensure that the content of a belief, derived from its functional role, is consistent with what the belief denotes or its satisfaction conditions. Thus, it provides no reliable formula for individuating belief. One-factor functional-role semantics fails because of the Twin Earth cases and two-factor functional-role semantics fails because it cannot provide a principled way in to keep its two factors coherent⁸ (Fodor 1987: 82).

⁸ Block (1993) does not really know what to make of this claim, since this is a problem for anyone constructing a theory of meaning who (like Fodor himself) recognizes the existence of functional roles and truth conditions. Moreover, he thinks that Fodor's claim that it is a major worry that the two factors may come apart is dubious, so he does not take this objection seriously. I propose not to either, and I will not discuss it in answering Fodor's objections below.

To sum up, functional-role semantics is supposedly the most promising way of justifying acceptance of the first step of the ur-argument for meaning holism. Fodor argues that the justification simply is not there because functional-role semantics cannot deliver what it promises. Later, I will argue that Fodor is wrong on these two counts: First, the ur-argument does not put us on a slippery slope to intentional anti-realism. Second, the problems that Fodor sees with functional-role semantics can be avoided. But the next order of business is consideration of the causal co-variance semantics that Fodor proposes.

2. Fodor's Theory of Content

Fodor argues that we should accept causal co-variance semantics because according to him it is the only semantic theory that is compatible with intentional realism and it is at least as plausible as any natural language-based theory of content. Since causal-covariance semantics fits with our common sense realistic account of psychology, it should win acceptance if it proves to be the only alternative to theories of content that entail intentional anti-realism. And supposing that Fodor's causal co-variance semantics is able to provide the most compelling theory of content, then it should constitute an independent argument for accepting the mentalese hypothesis.

Fodor argues that the content of an intentional state is a matter of lawful connections between the mind and the world, between tokenings of symbols in mentalese and the properties they express that reliably cause their tokenings. We derive meaning from the reliable causal co-variance between the presence of a property in the world and the tokening that denotes it. So 'X's mean X because all

and only Xs reliably cause tokenings of 'X.' It should go without saying that this theory is only plausible as applied to mentalese, since there is no reliable lawful connection between utterances and objects that they denote. There is no reliable connection between there being a mouse in my presence and my utterances of 'mouse.' For example, I may utter 'mouse' when there is not a mouse to be seen, or I may see a mouse and say nothing of it. If causal co-variance semantics proves to be the best account of intentional content, we would still need to devise a theory of meaning for natural language, perhaps with an account of its relation to mentalese. This would be a victory for mentalese over the supposition that we think in natural language, because it would undermine the intuitively plausible claim that thinking publicly and privately are one and the same, that talking is essentially 'thinking out loud.' But this is more easily conjectured than done, as there are obstacles for Fodor's theory to overcome before it even achieves plausibility.

The main difficulty for causal co-variance semantics is the disjunction problem. Inevitably, the mind will misrepresent objects in the world and mistakenly token an object as something that it is not. Suppose that an agent sees a little rat but misrepresents it and the little rat causes the tokening of a mouse representation in mentalese in the agent's mind. The presence of mice in standard conditions causes mouse-tokens, but it would also appear that sometimes rats also cause 'mouse-tokens.' Causal co-variance semantics trades on the assumption that content of intentional states is cashed out in terms of the properties that reliably cause them. This means that mouse-tokens do not just mean mouse, they mean mouse or small rat. We started with the assumption that the meaning of an intentional state is the product

of the lawful connection between its tokening and the thing that causes it. However, the possibility of error complicates matters, since we can be mistaken about more than just rodents. This is the disjunction problem, and for causal co-variance semantics to be viable it must explain the difference between the lawful connection between intentional contents that are actually disjunctive versus intentional states that are not disjunctive, but rather the product of error, like in the case of misrepresenting a rat as mouse.

Fodor's solution to the disjunction problem involves the recognition of the simple fact that the possibility of misrepresentation depends on the existence of lawful connections between objects and representations in the first place. A token misrepresentation is parasitic on the type-connection between a property and its lawful connection to a particular representation. To put it in the terms of the example above, in order to make sense of tokening 'That's a mouse' falsely, we must presuppose the existence of a nomological connection between 'mouse' and mouse-tokens. Thus, the meaning of 'mouse' depends only on the lawful connection between the presence of a mouse and the causation of a 'mouse' tokening. The fact that rats are sometimes misrepresented as mice is not necessary for tokens of 'mouse' to mean what they do, and rats would not produce false 'mouse' tokenings unless there were already a connection between tokenings of 'mouse' and mice. Thus, Fodor concludes that the causal connection between little rats and 'mouse' tokenings is asymmetrically dependent upon the causal connection between mice and 'mouse' tokenings (Fodor, 1987: 108). In summation:

'X' means X if and only if:

1. Xs cause 'X's' is a law (*all* Xs and *only* Xs reliably cause 'X's').

2. For all Ys not = Xs, if Ys qua Ys actually cause 'X's, then Y's causing 'X's is asymmetrically dependent on X's causing 'X's.
3. There are some non-X-caused 'X's.
4. Some 'X's are actually caused by Xs (This invokes the causal history of 'X' tokens as constitutive of the meaning of 'X,' but Fodor argues that it is necessary for condition (1) to be satisfied) (Adams, 2006).

Fodor believes that these conditions are necessary and sufficient for a solution to the disjunction problem. They allow us to tell apart the cases in which 'X' means X and the cases in which 'A v B' means either A or B like so. We can distinguish disjunctive tokenings by the fact that they are symmetrically dependent. The fact that A causes tokenings of 'A v B,' rather than just 'A,' is dependent on the fact that Bs are also lawfully connected to the tokening of 'A v B.' The representation of 'A v B' is symmetrically and lawfully connected to both experiences of A and experiences of B. In contrast, non-disjunctive properties are asymmetrically dependent, meaning that Xs are lawfully connected only to representations of 'X.' Ys may cause chance mistaken tokenings of 'X' but 'X' is never used to represent Ys.

Thus far we have focused on just the concern about the second condition for (1) to be a law, that only Xs must reliably cause tokenings of 'X's. How can we be sure that all Xs cause 'X's? Fodor's response is that our tokenings of 'X' refer to the property of Xness in the world, and to establish this fact, we must observe that *some* instances of 'X' must actually be caused by Xs (Fodor 1990: 121).⁹ Provided that we identify some instances of a property correctly, Fodor thinks we can generalize to all instances of a given property. In particular circumstances, a mouse can correctly cause 'mouse' tokens in virtue of the fact that it is similar in the relevant ways, by

⁹ This allows Fodor to rule out Putnam's twin cases by invoking the causal history of the entity in question. On this account of meaning, 'water' does not mean 'XYZ' because the causal history of our use of the term is attached to H₂O. Fodor worries that without this constraint there would be no way to distinguish the concept 'water' from 'twater.'

sharing the property of mousehood, with other mice that have also caused tokenings of mouse (Carruthers 1996: 99). From this we then have license to postulate a nomological relationship between mice and mouse-tokens.

3. Objections

Thus far I have considered two arguments of Fodor, one negative and one positive. Now I will provide some negative criticism of Fodor's arguments before offering a positive account of content in the next section. I will begin by voicing two problems I see with Fodor's theory of content. Then I will suggest a flaw in his criticism of holism, in order to clear some room for my proposal of a weakly holistic functional role theory of content.

Even if Fodor's theory of content is successful in providing sufficient conditions for an atomic, denotational theory of meaning, its complexity remains a strike against it. Fodor does not care that his sufficient conditions are very complex because he thinks that he is defending the only route to intentional realism. If we can provide a simpler way to the same destination, then Fodor cannot claim his theory of content has exclusive rights to common sense plausibility on the basis of preserving realism about propositional attitude ascriptions. Furthermore, Fodor's mistaken assumption that an atomistic theory of content is the only type of theory compatible with intentional realism causes him to deviate widely from an account of meaning that is true to the common sense practices of folk psychology.

First, Fodor's theory of content is out of keeping with our practice of propositional attitude ascriptions. It is a simple fact that our propositional attitude

ascriptions are not guided by asymmetric causal dependence (Carruthers 1996: 94). If someone calls a rat a mouse, thereby indicating the tokening of 'mouse' in her language of thought, even if I have read a lot of Fodor it is unlikely that I will suppose that her tokening and corresponding utterance of 'rat' is asymmetrically dependent on the causal co-variance between the presence of a mouse and the tokening of 'mouse.' This certainly does not reflect the common sense psychological talk that Fodor praises for being so successful. Suppose that we are faced with a real instance of the disjunction problem, and we wonder whether by 'mouse' this person means rat-or-mouse instead of just mouse. We would not query counterfactually, whether rats like this one would also have caused her tokenings of 'mouse' unless some 'mouse' tokens had already been lawfully caused by mice. We do not do this consciously anyway. The average person does not seek the satisfaction of conditions (1)-(4) above when they are trying to figure out what a person means. Instead, we try to figure out where this misrepresentation of mouse or representation of rat-or-mouse fit in with the rest of the agent's beliefs; we try to ascertain its functional role in her mental economy.

Carruthers' chemistry lab example best illustrates this point (1996: 95). Water and acid are can both be clear colourless liquids, and so they are sometimes easily confused. Suppose someone enters a lab and indicates that she thinks that a beaker of acid is 'water.' We may wonder, then, whether this is a mistaken tokening of 'water' or a correct tokening of 'water-or-acid.' In order to figure this out, we would not ask whether her tokenings of 'water' would have still been caused by acids if such tokens had also not already been caused by water. Neither would we ask if it were not for the

history of this person tokening 'water' in the presence of water, if she would have still tokened water in the presence of this acid. Instead, we would try to figure out how this agent's belief that 'this container of acid is water' fits in with the rest of her beliefs. So we would try to find out if this agent thinks that the liquid in question satisfies thirst, helps plants grow, and is good for doing your washing in. If the agent disagrees with any of these attributes, then she may very well mean 'water-or-acid' by her tokening of water. On the other hand, if the agent agrees, then we may very well show the agent that her judgment is false (thereby keeping her from harming herself by drinking it), by showing that the liquid actually disintegrates fabric rather than cleaning it.

This holistic practice of propositional attitude ascription fits far better with functional-role semantics than with Fodor's causal co-variance theory. When faced with the question of the content of a particular judgment, it makes much more sense to ascertain the content of the other beliefs relevant to the judgment, and then on those grounds make a judgment about what the agent means. We generally do not ask if this Y-judgment would have still been caused if it were not for other certain X-judgments being caused by X-objects. Causal asymmetric dependence has likely never occurred to any folk psychologist, and yet Fodor has gone to great lengths to construct a theory of meaning around it in the name of preserving realism about our folk psychological ascriptions. Of course, he can reply to this criticism that our practices of propositional attitude ascription are a poor reflection of the actual content of our propositional attitudes, but this is at the expense of his characterization of his position as one that defends common sense.

The second problem with Fodor's theory of content is that it "founders on the metaphysics of properties" (Carruthers 1996: 100). Even if we are sympathetic to his claim that there are real, mind-independent properties in the world that exist untouched by our own interests and systems of classification, surely we can only suppose that these properties encompass only the natural kind terms described by a completed science. This means that the meaning of many of our terms, which are largely dependent on our interests and do not pick out natural kinds, are left unaccounted for by Fodor's theory.

Carruthers cites 'sports' and 'spices' as examples of words that do not refer to a particular worldly property but instead classify things according to our own particular purposes. Fodor may reply that such terms can be cashed out in terms of disjunctives of world-properties in mentalese. So in this way our tokenings of 'sport' may refer to sports. Carruthers fundamental issue with this solution is that there is "no way of closing off the disjunction of properties falling within the extension of a term *except* by reference to variable human needs and purposes — which would violate Fodor's atomicity requirement" (1996: 101). We are inventing new sports all the time, and before we consider these activities to be 'sports' we do not check them against the list of disjunctive properties, the sum tokening of which, in mentalese, constitute our understanding of 'sport.' Rather, we check new sports against our own interests in sport. Similarly, before accepting a new sport as an 'Olympic sport,' we do not question the new sport matches up with the disjunctive list of properties in our language of thought that supposedly encompasses what a sport is, rather, we ask

whether a new sport is in keeping with the spirit and interests of the Olympic Games themselves.

If we combine the general moral of these two criticisms, that what matters when we individuate intentional content is where it fits in with the rest of an agent's mental economy and what interests are pertinent to the agent in question, then we have a prospective answer to Fodor's main criticism of two-factor functional-role semantics. Bilgrami says that we should make a distinction between two notions of content, local and aggregate (Bilgrami 1995: 335). Aggregate content is that which takes into account all beliefs relevant to the meaning of a given propositional attitude, whereas local content is that which accounts for only the beliefs that are relevant to a particular context of ascription. It may be possible, indeed it may be quite likely, that no two people share the same aggregate content because nobody shares all the same beliefs about a given term. However, this need not be a worry because the only content that matters is the content at the local level of propositional attitude ascription. These psychological explanations and generalizations are made possible through the selection of beliefs level that are obviously relevant to the locality in question from the aggregate. To illustrate this point, let us return to chemistry. Suppose that there is an earthling chemist in the kitchen with Twinnie from Twin Earth, who is completely ignorant of chemistry. At the aggregate level, we would say that the chemist and Twinnie have different beliefs about water, since the chemist likely has thoughts about H_2O and XYZ, whereas Twinnie is ignorant of such things. However, at the local level, in the kitchen, we can say that their belief selections about water (or even twater) can be the same. Both would agree and behave in a

manner that is consistent with thinking that what comes out of the tap is good for drinking and washing dishes in the sink. The essential chemical composition of water is inessential to the explanation of their behaviour in relation to the liquid, at least in the kitchen where no such questions are likely to arise.

4. Intentional Realism and Functional-role Semantics

The functional-role semantics that I want to advocate does not result in the rampant holism that worries Fodor, but rather the *weak* holism¹⁰ that Carruthers and Bilgrami believe has great utility for common sense psychological explanations. The way we accomplish this is by refusing to accept every connection to a propositional attitude as being relevant to the individuation of an intentional content. Functional-role semantics requires only that we understand the content of a mental state in terms of its *potential* causal interactions with bodily stimuli, with other mental states, and with behaviour (Carruthers 1996: 116). Hence, we do not necessarily need to worry about incorporating related actualized states into our individuation of an intentional state. To borrow an example from functionalism, the fact that I desire to be brave and not show my pain, whereas you desire to advertise it in a plea for sympathy does not change the fact that our pain-states are the same. Moreover, if *all* of our mental states were the same, then we could suppose that the effects of our pain-states would be the same.

In this way, we can understand how functional-role semantics individuates the content of an intentional state not in terms of the sum-total of its epistemic liaisons,

¹⁰ For an explanation of how weakly holistic semantics is compatible with compositionality, see Block's paper "Holism, Hyper-analyticity and Hyper-compositionality" (1993).

but rather in terms of the closed set of conditionals about what the subject would do if they had other specified propositional attitudes (Carruthers 1996: 116). These hypothetical suppositions can be generalized over a variety of individuals with a variety of different actualized beliefs and desires, so this brand of functional-role semantics poses no problem to the scientific approach to folk psychology that we hold dear. What determines this closed set of conditionals will depend not on the aggregate of beliefs relevant to a certain intentional state, which could lead us down Fodor's slippery slope to the assumption that all beliefs determine its content. Rather, the interests of the agent understood within the locality of explanation should provide a limit to the relevant intentional states, such that we may individuate the content of a particular propositional attitude.

In this chapter, I have described Fodor's criticisms of theories of content based in natural language, particularly functional-semantics, and offered an account of his proposed semantics for mentalese. I have found Fodor's arguments wanting and criticized them accordingly. Finally, I have offered an account of how functional-role semantics is compatible with intentional realism and more in keeping with our common sense folk psychological beliefs than Fodor's own theory.

Conclusion

In the three chapters of this thesis I have set out to accomplish three different goals in order to reach the final conclusion that natural language should be afforded primacy over mentalese in the study of the nature of the mind. In Chapter One, I argued that all the phenomena that of which the mentalese hypothesis purportedly provides the best explanation are better explained by supposing that natural language performs the explanatory role. This is clear if we account for considerations of parsimony and fruitfulness. Positing and testing for one system of representation is simpler than positing two, especially since the second system of representation posited by the mentalese hypothesis escapes our conscious awareness and has thus far eluded any discovery by scientific investigation. In Chapter Two, I argued that introspection gives us good reason to think that insofar as our conscious propositional thought is reflexive, it requires natural language as its medium. In Chapter Three, I concluded with a criticism of Fodor's claim that the only viable semantics for an intentional realist about mental states requires that we accept the mentalese hypothesis. Together these chapters show that we have good reason to investigate the function of natural language in cognition.

BIBLIOGRAPHY

Adams, Fred. "Fodor's Asymmetrical Causal Theory of Meaning" from <URL=<http://host.uniroma3.it/progetti/kant/field/asd.htm>> Retrieved April 2006.

Akeel, Bilgrami. "Why Holism is Harmless and Necessary" Vol. 32 Supplement Philosophical Perspectives (1998).

Abbott, Barbara, (1995) "Thinking without English" in Behavior and Philosophy, Vol. 23, pp. 47-52, URL=< <http://members.aol.com/lshauser/tweng.html>> Retrieved Feb. 2007

Akeel, Bilgrami. "Review of *A Theory of Content and Other Essays.; Holism: A Shopper's Guide*" in *The Journal of Philosophy* Vol. 92 No. 6 (June 1995).

Amaral, Pedro Ed., *The Lectures of Wilfrid Sellars on the Metaphysics of Epistemology*, California: Ridgeview Press, 1989.

Aydede, Murat, "The Language of Thought Hypothesis", *The Stanford Encyclopedia of Philosophy* (Fall 2004 Edition), Edward N. Zalta (ed.), URL=<<http://plato.stanford.edu/archives/fall2004/entries/language-thought/>>, Retrieved April 2006.

Bermudez, Jose Luis, *Thinking Without Words*. New York: Oxford University Press, 2003.

---. "Nonconceptual mental content", *The Stanford Encyclopedia of Philosophy* (Jan 21 2003), Edward Zalta (ed.), URL=<<http://plato.stanford.edu/entries/content-nonconceptual/>>, Retrieved December 2006.

---. "Naturalism and Conceptual Norms" *The Philosophical Quarterly*, Vol. 49, no. 194 (1999). pp. 77-85.

---. "Nonconceptual content: From perceptual experience to subpersonal computational states." *Mind and Language* Vol. 10 (1995), pp. 333-369.

Block, Ned. "Advertisement for a Semantics for Psychology," in *Midwest Studies in Philosophy*, Vol. 10. (1986), pp. 615-678.

---. "Holism, Hyper-Analyticity and Hyper-Compositionality" in *Philosophical Issues* Vol. 3, Science and Knowledge (1993), pp. 37-72.

Carruthers, Peter, *Language, thought and consciousness*. Cambridge: Cambridge University Press, 1996.

---. "The Cognitive Functions of Language" in *Behaviour and Brain Sciences*, Vol. 25 no. 2 (2002)
 URL=<<http://www.philosophy.umd.edu/Faculty/pcarruthers/Cognitive-language.htm>> Retrieved Dec. 2006.

Carruthers, Peter and Jill Boucher Eds. *Language and Thought: Interdisciplinary themes*. Cambridge: Cambridge University Press, 1998.

Clark, Andy. *Mindware*. New York: Oxford University Press, 2001.

---. "Magic Words: How Language Augments Human Computation" in *Language and Thought: Interdisciplinary themes*. Eds. Peter Carruthers and Jill Boucher, Cambridge: Cambridge University Press, 1998.

Cole, David. (1997). "Hearing Yourself Think: natural language, inner speech and thought", URL=<<http://www.d.umn.edu/~dcole/hearthot.htm>>. Retrieved January 2007.

Curtiss, Susan. "Abnormal language acquisition and the modularity of language" in Newmeyer, F. J. (Ed.), *Linguistics: The Cambridge Survey. Volume II Linguistic Theory: Extensions and Implications* (pp. 96-116). Cambridge: Cambridge University Press, 1988.

---. *Genie: A Psycholinguistic Study of a Modern-day "Wild Child"*. New York: Academic Press, 1977.

Davidson, Donald. "Thought and Talk" in *Inquiries into Truth and Interpretation*. New York: Oxford University Press, 1984, pp. 155-171.

Davidson, Donald. *Subjective, Intersubjective, Objective*. New York: Oxford University Press, 2001.

Dummett, Michael "The Origins of Analytical Philosophy" pp. 335-360 in *Understanding and Sense II*. Ed. Christopher Peacocke, Dartmouth: Dartmouth Publishing, 1993.

Fodor, Jerry. *A Theory of Content and Other Essays*. Cambridge: MIT Press, 1990.

---. "Do We Think in Mentalese? Remarks on Some Arguments of Peter Carruthers" in *In Critical Condition*. Cambridge, MIT Press 1998. pp. 63-74.

---. *The Language of Thought*. New York: Crowell Company, 1975.

---. "Propositional Attitudes," (1978) reprinted in *Philosophy of Mind*, Ed. Chalmers, Oxford University Press 2002, pp. 542-556.

---. *Psychosemantics*. Cambridge: MIT Press, 1987.

---. Review of Bermudez's *Thinking without Words*. (2003) <Url=<http://consc.net/online2.html#language>> Retrieved Jan 2007.

Fodor, Jerry and Ernest Lepore. *Holism: A shopper's guide*. Cambridge: Blackwell, 1992.

Gauker, Christopher. *Thinking Out Loud: An Essay on the Relation between Thought and Language*. Princeton University Press, 1994.

Hauser, Larry. "Doing without Mentalese" in *Behavior and Philosophy*, Vol. 23 (1995), pp. 42-47, URL=< <http://members.aol.com/lshauser/dwment.html>> Retrieved Feb. 2006.

Hume, David. *A Treatise of Human Nature*, 1739.

Hurlburt, R. T., and Heavey, C. L.. *Sampling Inner Experience with Disturbed Affect*, Plenum Press, 1993.

---. *Sampling Normal and Schizophrenic Inner Experience*. Plenum Press, 1990.

---. "Telling what we know: describing inner experience." *Trends in Cognitive Sciences*, Vol. 5 (2001), 400-403.

Jackendoff, Ray. "Chapter 8 Epilogue: How Language Helps Us Think" *The Architecture of the Language Faculty*. MIT Press, 1997, pp. 179-194.

Johnson-Laird, P., and R. M. J. Byrne. *Deduction*. New Jersey: Erlbaum, 1991.

Kaye, Lawrence. (1998). "Another Linguistic Turn? Review of *Language, Thought and Consciousness*", URL=<<http://psyche.cs.monash.edu.au/v4/psyche-4-02-kaye.html>> Retrieved 17 Dec 2006.

Kaye, Lawrence J., "The Languages of Thought" in *Philosophy of Science*, Vol. 62, No. 1 (1995), pp. 92-110.

Keller, Helen. *The World I Live In*. London: Hodder and Stoughton 1909.

Laurence, Stephen and Eric Margolis. "Regress Arguments Against the Language of Thought," in *Analysis*, Vol. 57, No. 1 (1997).

Locke, John. *An Essay Concerning Human Understanding*, 1690.

Pinker, Steven. *The Language Instinct*. Penguin Books, 1994.

Popper, Karl. "7. Simplicity", *The Logic of Scientific Discovery*, 2nd edition, London: Routledge, 1992, pp. 121-132.

- Putnam, Hillary. Putnam, Hilary "The Meaning of 'Meaning' ", reprinted in his *Mind, Language, and Reality*. Cambridge: Cambridge University Press, 1975, pp. 215-271.
- Russell, Bertrand. *The Analysis of Mind*. London: Allen and Unwin, 1921.
- Silby, Brent. *Revealing the Language of Thought*, 2000, URL=< <http://www.def-logic.com/articles/RevealLanguageOfThought.html>> Retrieved Mar. 2007.
- Machery, Edouard. "You Don't Know How You Think: Introspection and the Language of Thought," *The British Journal of Philosophy of Science*, Vol. 56 (2005), pp. 469-485.
- Nisbett, Richard and Wilson, Timothy, "Telling More than we can Know", *Psychological Review*, Vol. 84 (1977).
- Rosenthal, D. M. "Two Concepts of Consciousness" in *The Nature of Mind*, Ed. David Rosenthal. New York: Oxford University Press, 1991, pp. 462-477.
- Sellars, Wilfrid. "Empiricism and the Philosophy of Mind" (1956) from <URL=<http://www.ditext.com/sellars/epm.html>> Retrieved Sept. 2006.
- Wittgenstein, Ludwig. *Philosophical Investigations*. Trans. G.E.M. Anscombe, Oxford: Blackwell 1953.