Thoughts Animals Can Think: Attributing Beliefs and Describing Content

by

Kevin J. Krein

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Graduate Department of Philosophy

University of Toronto

© Copyright by Kevin J. Krein 2001
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

0-612-63744-1
Thoughts Animals Can Think: Attributing Beliefs and Describing Content

Doctor of Philosophy 2001
Kevin J. Krein
Graduate Department of Philosophy
University of Toronto

Abstract

One of the problems that arises in attempts to attribute thoughts to non-human animals is that it is difficult to find sentences of human language that accurately express the content of those thoughts. I argue that semantic theories that do not depend on an understanding of propositional content as being linguistic in nature can provide a helpful perspective from which to approach questions about the content of the beliefs of non-language-using animals. In particular, I consider Robert Stalnaker’s possible worlds semantics and the situation semantics of Jon Barwise and John Perry. I argue that both of these theories provide ways of understanding the content of beliefs on which it is far more plausible that non-human animals have beliefs than theories that take propositions to be linguistically structured. I further argue, however, that Barwise and Perry’s account is ultimately more helpful than Stalnaker’s and avoids some significant problems faced by possible worlds theories. In my approach to these topics, I consider the plausibility, on various influential theories of mind, of the idea that animals have beliefs and I provide an exposition of Jonathan Bennett’s account of what types of animal behavior justify the attribution of beliefs. I conclude by arguing that the combined work of Bennett, Stalnaker, and Barwise and Perry can be used to address a line of reasoning I refer to as the specificity argument. According to
the specificity argument, if an animal cannot express its beliefs in language, it is impossible to attribute content to any belief one might attribute to it with enough precision to make sense of claims that the animal has one belief rather than another. It is then concluded that non-language-using animals do not have beliefs. I use the work discussed above to show that the specificity argument does not, by itself, show that non-human animals do not have beliefs. As well, the approach to understanding content that I consider will be helpful in considerations of the content of non-language users in general, and is a promising first step in the assessment of many arguments against the claim that animals have beliefs.
Acknowledgements

Completing this thesis required slightly more time than most of the people who are close to me expected. In addition, completing it required much of my energy and attention. I owe much to many people who supported and encouraged me. I could not have finished without their patience and help.

I would like to thank my supervisor, Ronald de Sousa, who devoted his time, energy and expertise to helping me finish. Professor Bernard Katz offered encouragement at times when I needed it. I doubt that either one of them know the extent to which they have inspired me as philosophers.

I owe a great debt to Abigail Levin for all of her help. Over the past few years, she generously read many drafts of every part of this thesis and spent hours discussing the philosophical ideas it contains. She often had more confidence in me, and the work I was doing, than I did; I cannot thank her enough for that. The help and emotional support she provided were invaluable.

As well, I would like to thank Tony Kostroman and Lucia Dow. The three of us began graduate school together and the two of them continue to deeply influence my thinking and approach to philosophical issues.

Finally, I wish I could express the gratitude I feel for the unwavering support, encouragement, and love my parents have provided. Without these things, it is unlikely that I would have begun graduate work at all and it is certainly true that I would not have been able to finish it.
# Table of Contents

Chapter One:  
Might Animals Have Beliefs?  

Chapter Two:  
Teleological Explanations and the Attribution of Beliefs  

Chapter Three:  
Stalnaker on Possible Worlds and Beliefs

Chapter Four:  
Problems With Possible Worlds

Chapter Five:  
Situation Semantics

Chapter Six:  
The Specificity Problem

Works Cited
Chapter One
Might Animals Have Beliefs?

The question of whether thought depends on language is, and has been, a key issue in the philosophy of mind. This thesis contributes to that debate by providing a close examination of issues relating to problems concerning the accuracy of attributions of beliefs, and more generally thoughts, to non-language-using animals. The bulk of the thesis is a defense of the claim that the content of the beliefs of non-language-users can be accurately known and represented. It is further argued that once the content of the beliefs of non-language users is accurately represented, some of the philosophical problems that arise with the attempt to attribute thoughts to animals can be dealt with.

This assumes that we find it at least possible that there are beings that think, but do not have language. Following these introductory remarks and a short explanation of my terminology, I will, in the first two chapters, provide an explanation of why one might think that animals have thoughts. To this end, I begin by arguing that some of the theories of mind that are currently the most well received among philosophers are consistent with the idea that animals think. I have chosen three influential contemporary philosophers of mind: John Searle, Jerry Fodor, and Daniel Dennett, whose positions I will describe briefly. I point to some criticisms of these positions, but do so in the spirit of clarifying the issues they confront. My purpose is not to provide in depth comparisons and criticisms of their theories of mind, but instead, to present their positions in enough detail to illustrate that the
fundamental ideas of each present no reasons not to attribute thoughts to non-human animals.

I will not attempt to argue for one theory of mind over another in this thesis. That is, I will not attempt to determine what the nature of thought is. Instead, I will argue that the claims that I make concerning the attribution and content of the beliefs of animals are consistent with generally accepted theories of mind. As well, I will not attempt to incorporate theories of consciousness in this thesis. Consciousness, it seems to many philosophers, is, to some degree at least, separable from intentionality.¹ If so, then it is possible that some animals are conscious — can experience pain, pleasure and other sensations — but neither think nor have beliefs. I will not attempt to address the question of whether it is possible that one has beliefs but is not conscious.

In Chapter Two I will discuss Jonathan Bennett’s work on the question of what types of behavior of a non-language-using animal justifies the attribution of belief to it. Chapters One and Two establish a starting point for discussion of the difficulties in accurately attributing contentful beliefs to non-language-users. This discussion occupies the rest of the thesis.

Chapters Three and Four focus on Robert Stalnaker’s theory of propositional content and how it can be used to represent the beliefs of non-human animals. In Chapter Three I will begin to explore the question of how the contents of the beliefs of non-language-using animals can be represented by introducing Stalnaker’s possible

¹My use of ‘consciousness’ should be understood to follow Nagel’s position that, if a being is conscious then there is something that it is like to be that being, as presented in “What is it Like to Be a Bat?” The Philosophical Review LXXXIII, 4 (October 1974): 435-50.
worlds semantics. I argue further that such an account can be combined with Bennett's work in order to determine if any non-human animals have beliefs. In Chapter Four I address some common criticisms of possible worlds accounts of content.

While I attempt to address criticisms of possible worlds accounts of content in Chapter Four, there are, I argue, problems with Stalnaker's account that are dealt with better by the situation semantics of Jon Barwise and John Perry. In Chapter Five, I discuss these problems and argue that situation semantics will provide the same benefits as possible worlds semantics in representing animal beliefs and will avoid many of the theoretical difficulties faced by possible worlds semantics.

In Chapter Six, I argue that the adoption of either of the non-linguistic theories of content, possible worlds semantics or situation semantics, will help in dealing with what I refer to as the specificity argument. It is often claimed, against those who hold that non-language-using animals have beliefs, that if an animal does not use natural language, it is impossible to attribute beliefs to it with enough precision to make sense of the claim that such a creature has one belief rather than another. It is then concluded from this, that non-human animals do not have beliefs. I deal, in particular, with versions of the specificity argument that have been presented by Donald Davidson and Stephen Stich. I argue that with a combination of the work of Bennett and the adoption of a non-linguistic account of content, it can be demonstrated that the specificity argument does not, by itself, show that non-human animals do not have beliefs.
Before moving to the main topic of this chapter, I will provide some preliminary clarification of my use of various terms. I will not offer any radical definitions that are at odds with the standard philosophical uses of most terms. Still, clarification of my own use of terms, especially those commonly found in descriptions of the mental, will be helpful.

‘Animal’ is, of course, a key term in this thesis. For the sake of simplicity I use ‘animal’ to refer to non-human animals. This does not reflect any philosophical position according to which humans are not understood to be a type of animal.

I will often refer to animals as non-language-users. I use ‘language’ in Noam Chomsky’s sense of the term to refer to systems of symbols that have both semantics and syntax in which the symbols can be combined and recombined in different ways to generate an infinite number of meaningful sentences.\(^2\) Chomsky argues that these features are found only in human languages and that animal communication systems are not languages. He explains that every known system of animal communication uses one of two basic principles:

Either it consists of a fixed, finite number of signals, each associated with a specific range of behavior or emotional state...or it makes use of a fixed, finite number of linguistic dimensions, each of which is associated with a particular nonlinguistic dimension in such a way that selection of a point along the linguistic dimension determines and signals a point along the associated nonlinguistic dimension (Chomsky 61).

---

The first type of system Chomsky refers to is very straightforward. Examples of this type of communicative behavior are found everywhere — for example, when a wolf or dog tucks its tail between its legs, signaling submission, or when a beaver slaps its tail on the water, alerting other beavers in the area to danger. As an example of the second type of communication, Chomsky cites the European robin, which signals its readiness to defend its territory by alternating between high and low pitched signals— "the higher the rate of alternation, the greater the intention to defend the territory" (Chomsky 61). In the communication system of the European robin there is a continuum along which an indefinite variety of symbols are available; at least in principle, each spot on the continuum represents a different level of intensity of readiness to defend territory. Chomsky's point is that, while communication of some sort takes place, "...the mechanism and principle...are entirely different from those employed by human language to express indefinitely many new thoughts, intentions, feelings and so on" (Chomsky 31). When a sentence is constructed in human language, he argues, the speaker does not choose a signal from a continuum, nor pick out a single signal that corresponds to a single behavior or object in the world. Meaning depends, among other things, on how, and in what order, symbols are combined. Chomsky maintains that although animal communication systems have much in common with human gestural systems, they have little in common with actual human language (Chomsky 62).

---

This is not to deny that many animal species have communication systems which are rich in usefulness and central to the lives of members of those species. It is only to assert that there are very important differences between human language and the communication systems of wild animals. The point of this thesis is to consider whether it is necessary, in order for an animal to have beliefs with specifiable content, that it use language, in Chomsky’s sense. I argue that it is not.

It is worth noting that debates over whether primate sign language constitutes actual language are not as central to this topic one might assume. The question I am trying to answer is not whether or not it is only humans that have thoughts with specifiable content, but rather, is it possible to have such thoughts without being a language-user? Chimpanzees trained in the use of American Sign Language, for example, present borderline cases of language use — while chimpanzees use of sign language is similar to human language in many respects, it is dissimilar in important respects also. If the elements of syntax are present at all in chimpanzee sign language, the degree to which they are is extremely restricted. Depending on what types of conclusions we reach concerning whether or not chimpanzees that have been taught sign language by humans have minds, the combination of facts concerning their language abilities and facts concerning whether they think, could provide valuable insight into which characteristics of language are of critical importance in the ability to think. This is of course, very important. However, if what we are concerned with,

---

4 For an account of the most successful project of this sort see Sue Savage-Rumbaugh and Roger Lewin, *Kanzi: The Ape at the Brink of the Human Mind* (New York: John Wiley and Sons Inc., 1994).
5 For example, Terrence Deacon’s discussion of primate language learning seems to shed light on aspects of language and mind that help to pinpoint and explain important differences between animals
first and foremost, is the question of whether language is necessary for having
thoughts with specifiable content, the most useful examples will most likely be those
animals for which it cannot even be argued whether they use a language of the sort
humans use, yet seem to be thinking all the same.

Turning to properties of mental states, one characteristic feature that is central
to philosophical discussions of belief and thought is intentionality. ‘Intentionality’, as
I use it, refers to the quality of directedness, or, as Brentano put it, “reference to a
content” or “direction toward an object.” According to Brentano, intentionality is the
mark of the mental — all and only mental states are intentional.

Much of the work in recent philosophy of mind has been on the question of
where such intentionality could come from, attempting to explain how it is that some
things in the universe (most often mental states) can be *about* other things. Searle,
Dennett, and Fodor provide differing accounts of the origins of intentionality, which I
will examine shortly. While it may not be the case that all mental states are
intentional,\(^7\) we can see that, at least, beliefs are, since all beliefs are *about*
something: the belief that ‘Mt. McKinley is 20,320 ft. high’ is about Mt. McKinley
and its height and is in this sense intentional. It makes little sense to speak of a belief
that is not about anything.

---

\(^1\)Deacon’s discussion is found in his, *The Symbolic Species*, (New York: W.W. Norton and

\(^6\)Franz Brentano, *Psychology from an Empirical Standpoint*, trans. A. C. Rancurello, D.B. Terrell, and

\(^7\) John Searle, for example, argues that emotional states such as depression are not always intentional
— one may be depressed about nothing in particular. John Searle, *Intentionality* (Cambridge, UK:
A useful and popular way of describing intentional mental states has been to say that each has a semantic or propositional content that is adopted in a particular way. I can believe or hope or fear that 'Mt. McKinley is 20,320 ft. high', in each case adopting a different attitude toward the same proposition.

In addition to mental phenomena, there seem to be many non-mental entities, like books and the sentences in them (and representations such as pictures, graphs and maps) that are intentional. This type of intentionality has been termed derived intentionality. The intentionality of such artifacts, it is often argued, is derived from their being created and used by beings that possess original or intrinsic intentionality. Sentences get their meaning by virtue of the role they play in languages used by human beings. If animals have intentional states, the intentionality of those states must be original intentionality.

In my discussions of intentional states, I will refer mainly to beliefs, with the assumption that one who is capable of believing is capable of having other types of intentional states as well. It does not follow from this that one who is capable of having beliefs is capable of all other intentional states. It may be that there are some propositional attitudes that require more sophisticated relationships to their objects than does belief and there is no reason to assume that one who is capable of believing is capable of all other more sophisticated relations. Still, I will work with the assumption that once we have an understanding of the contents of beliefs, we are

---

closer to having an understanding of the content of most other intentional states as well.

I have used 'belief' and 'thought' together and will continue to do so. The relation between these terms needs some clarification. As I will use the term, to have a thought is to be in an intentional state with a specifiable content. Beliefs then, are perhaps the most basic kind of thoughts and are required for any more sophisticated thoughts to take place. If one hopes, perhaps during a geography test, that 'Mt. McKinley is 20,320ft. tall,' he or she must also believe that Mt. McKinley is a mountain, that a foot is a unit of measurement, and other things which, if lacking, would make it impossible to attribute any content to the hope the person is said to have. Hence, if a being can have thoughts of any type, it is difficult to imagine that that being could be incapable of having beliefs.

I will use 'thinking' to refer only to mental processing that involves intentional states that have specifiable content. According to this use of the term, if one is thinking, he or she is having thoughts. My use of the term is more stringent than that of some philosophers who also use 'thinking' to refer to mental processes that are intentional — in the sense of having the quality of directedness described above — but do not have specifiable contents. Such philosophers argue that there are states which, while without their own specifiable content, have intentional objects that can be identified. Norman Malcolm, for instance, maintains that there is a
distinction between the claims that a being *thinks* and that a being *has thoughts*. Malcolm argues that while animals are capable of being in intentional states, they are not able to harness propositional contents, as do language-speaking humans. In this case, having a thought is seen to involve being in some kind of relation to a proposition while thinking can be understood to be reacting to one's surrounding while doing a fair amount of mental processing. According to views such as Malcolm's, an animal may think such and such, but not have the thought that such and such, which involves holding a proposition in one's mind. Malcolm maintains that when we say of animals that they think, "...we employ a verb that, grammatically, takes a propositional expression as object, without meaning that as a matter of psychological fact the animal thought of a proposition or thought *via a proposition*" (Malcolm 50). Malcolm does not explicitly say that the mental states of non-language-using animals do not have specifiable content. He does say however, that animals think, but do not do so via propositions. I will discuss Malcolm's position in detail in Chapter Six and argue against it. My use of 'thinking' throughout all other chapters is synonymous with 'having a thought'.

'Intentionality', as I have been using it, applies to some mental states. I will however, refer also to intentional *actions* or *behaviors*, meaning those actions which are intended or done on purpose. Doing something on purpose presupposes fairly

---

10 As stated above, I will use "thinking" only to refer to mental processes in which thoughts play a role. In claiming that thoughts have specifiable content, however, I do not mean to claim that the content of thoughts is completely determinate. In Chapter Six, I will support the claim that the content of thoughts, whether belonging to a non-language-using animal or a human, while specifiable to different degrees, will in all cases be somewhat indeterminate.
sophisticated mental activity. Intentional behavior is only possible for a being that is capable of having mental states that are intentional (in Brentano's sense) to motivate them. It makes little sense to speak of a being doing x intentionally or on purpose if that being cannot also be said to believe that it is doing x. To say that an action is intended then, or intentional, is to say that it is the product of intentional states and is the behavior of a being capable of such states. I will also speak of purposive behavior, or behavior for which a teleological explanation is relevant. I will argue that such behavior is not dependent on the ability to have beliefs. I will have more to say on this in Chapter Two.

These terms are, then, related to each other in the following ways. It may be possible that some creatures have intentional states, but do not have beliefs or thoughts. A belief, or thought, is a variety of intentional state that involves a specifiable content. Only beings that are capable of having intentional states with such content are capable of intending or acting on purpose or for a purpose. But, some purposive behavior is accomplished by beings that do not have the capacity to have beliefs.

Throughout this thesis, I will be making reference to a variety of approaches to the mind. I have discussed how I will use each of the terms mentioned above but will no doubt run into variance in usage by other philosophers. If one of the terms discussed above is used in a way that is different from that given above, I will make note of it.
I now turn to an examination of three prominent positions on the nature of intentional states. I will argue that each of these positions is in no obvious way inconsistent with the idea that language users can think. I begin with a discussion of the theory of mind supported by John Searle, then turn to that of Jerry Fodor, and finally that of Daniel Dennett. As stated earlier, this discussion, combined with Chapter Two's presentation of Jonathan Bennett's arguments concerning when it is appropriate to attribute beliefs, serve to establish a starting point from which to discuss further questions pertaining to the representation of the content of animal beliefs.

According to John Searle's position, which he has called 'biological naturalism,' intentionality does not depend on language. According to Searle, intentionality,\(^\text{11}\) which he describes as "...that property of many mental states and events by which they are directed at or about or of objects and states of affairs in the world," (Searle 1983, 1) arises as a natural product of certain biological systems. As products of biological systems, Searle claims that:

> ...mental states are as real as any other biological phenomena, as real as lactation, photosynthesis, mitosis, or digestion. Like these other biological phenomena, mental states are caused by biological phenomena and in turn cause other biological phenomena (Searle 1983, 264).

On this picture, intentional states are caused by the brain and realized within it. Hence, organisms with brains capable of producing intentionality can be in states that are directed toward or are about things in the world. Searle makes similar claims

---

\(^{11}\) Searle capitalizes 'intentionality' to demarcate his use of it as a technical philosophical use from its standard use of referring to an intention to do a particular thing or act in a particular way. I will not follow his convention except when quoting him directly.
concerning consciousness: "Consciousness, in short, is a certain biological feature of certain animal brains. It is caused by neurobiological processes and is as much a part of the natural biological order as any other biological features such as photosynthesis."12

One of Searle's most influential arguments in favor of his claim that intentionality is a property of biological states is his Chinese room argument.13 The argument is based on a thought experiment in which one is asked to consider that she is placed in a room in which there are a number of baskets containing Chinese characters and a set of rules and that she does not understand Chinese. People can pass sets of Chinese characters through a slot to the person in the room who then follows the rules and passes a set of characters, arranged according to the rulebook, back through a slot. The sets of characters people pass through the slots can be questions in Chinese and the sets passed back, if the rulebook is good enough, can be appropriate answers in Chinese. Searle points out that no matter how good the person gets at following the rules for manipulating the symbols, the person is not speaking Chinese. The person in the room may not even know that the symbols are Chinese characters or that the sets of symbols being passed through the slot are questions for which the rulebook determines answers.

The rules in the book are rules of syntax; they are rules concerning how symbols should be combined and do not tell the person in the room what the meaning of any of the symbols are. Searle argues that for the same reason the person in the

room cannot be said to understand Chinese, a computer program could never be a mind:

...a computer has syntax, but no semantics. The whole point of the parable of the Chinese room is to remind us of a fact we knew all along. Understanding a language, or indeed, having mental states at all, involves more than just having a bunch of formal symbols. And, a digital computer...cannot have more than just formal symbols...\(^\text{14}\)

Searle claims in more recent work that he does not argue that machines could never think and that it is a misunderstanding to take this to be his position: "The brain is a machine, a biological machine, and it can think. Therefore, at least some machines can think, and for all we know it might be possible to build artificial brains that can also think."\(^\text{15}\) Even so, it is Searle's position that intentionality as we know it, or, at least, original intentionality, is a product of biological systems.

Searle is notoriously unable to give an account of how a physical process might produce a conscious or intentional state. However, on Searle's account, this is a job for neuroscience rather than philosophy. While a philosopher might be able to say some interesting things about the ontological status of digestion, no one would expect that the best person to ask about the causal processes involved in the breakdown of food in the small intestine would be a philosopher. The same is true, Searle claims, in the philosophy of mind. Searle explicates his conception of the study of the biological nature of intentionality by way of analogy with the study of physical processes that occur along with visual experiences. He points out that the study of vision includes investigation of the workings of the retina-photoreceptor cells (ganglion, amacrine,


bipolar...) as well as the tracing of signals through the optic nerve, to the geniculate nucleus to the cells and the striate cortex which diffuse signals to specialized cells within the visual cortex, and so on. He says the following of this type of scientific description:

Notice that this story is a causal account, it tells us how the visual experience is caused by the firing of a vast number of neurons at literally millions of synapses. But where, then, is the visual experience in this account? It is right there in the brain where these processes have been going on. That is, visual experience is caused by the functioning of the brain in response to the external optical stimulation of the visual system, but it is also realized in the structure of the brain (Searle 1983, 267).

It seems clear that if we are ever going to understand visual experiences, we will need to have a lot of knowledge of how eyes and brains work.

One might criticize Searle for taking a description that names areas of neural activation and tells us how the eye works to pass signals through the optical nerve to the brain, as telling us how visual experience is caused. As Searle describes this type of study, the really mysterious parts—the parts about how something could be a conscious experience and how intentionality fits in — is left out completely. Searle's response to this kind of criticism is that the mind-body problem is similar to water molecule-wetness problem (our descriptions of the interaction of water molecules at given temperatures do little to explain the wetness of water at room temperature.) Of this Searle says:

If one knew the principles on which the system of H₂O molecules worked, one could infer that it was in a liquid state by observing the movement of the molecules, but similarly if one knew the principles on which the brain worked one could infer that it was in a state of thirst or having a visual experience (Searle 1983, 268).
Searle may be correct if he is claiming that, in each of the two cases, certain states in which the system may be found can be correlated with other properties of the system. However, discovering that events of a certain type can be correlated with events of a different type can be useful without explaining very much; one might be able to predict when it is going to rain by looking at the clouds, yet lack an understanding of the principles which govern the weather. Even if neuroscience could provide correlate signals in the brain with intentional states, it is not at all clear that this approach would yield an understanding of the mind.

There is a better interpretation. Searle could be understood to be claiming that neuroscience will lead to an understanding of the principles that produce intentionality. On this reading, his claim concerning the powers of neuroscience makes more sense, but does not seem supportable unless some description of how it might be that properties of intentionality could be produced as biological features of a physical organ can be given. Searle’s theory, and the theories that similarly take intentionality to be a biological property of functioning brains, ultimately depend on the possibility of scientific theorists finding and explaining the mechanisms which such theories claim exist. One way to assess Searle’s position is to examine the likelihood of neuroscience explaining how it is that brains produce intentionality.

---

16 I say “may be” because Searle’s position is entirely dependent on the idea that content can be ascribed independently of external factors. It has been argued by Hilary Putnam, Tyler Burge, and many others that factors external to individuals, to which thoughts are ascribed, play a role in determining the content of those thoughts. This, however, is not the place to take up that debate.

17 A recent attempt at giving a scientific explanation of mind is found in A.G. Caim-Smith’s, *Evolving the Mind*. This is the kind of work in which the explanations needed to fill in the gaps in Searle’s picture could be found. However, in this book at least, there seems to be no need for a notion
It seems fairly obvious that Searle’s position, in taking intentionality to be a biological property of functioning brains, is compatible with the claim that thinking does not require language. As Searle explains it, all that is needed in order for a creature to have intentional states, is the right sort of brain. Once we understand how it is that brains produce intentionality, we will not have to observe a creature’s behavior, environment, or interactions with other beings, in order to determine whether or not it has intentional states. We could look directly at any brain and find this out. In Searle’s opinion, intentionality is in no way dependent on language and in support of this claim he says, “...it seems to me obvious that infants and many animals that do not in any ordinary sense have a language or perform speech acts nonetheless have Intentional states” (Searle 1983, 5). He goes on to make the rather strong statement that, “Only someone in the grip of a philosophical theory would deny that small babies can literally be said to want milk and that dogs want to be let

of intentionality that is as robust as Searle’s. Instead, Cairns-Smith claims that he follows Victor Serebriakoff in seeing intelligence as “...'optimizing behavior in the light of information' without prejudice as to how this is achieved, whether consciously or automatically, whether it is a lot or a little.” Cairns-Smith goes on to say:

Even my refrigerator, which knows when it has to switch its cooling motor on and off, is a tiny bit intelligent on that score: about as much so as an allosteric enzyme which will switch off when enough of its product has been made and on again when more is needed. And whole cells are bright. They are able to cope wonderfully, to respond appropriately to threats and opportunities, to keep conditions inside right for near optimum performance...(Cairns-Smith 91).

As will become clear, I am not advocating a position such as Cairns-Smith’s. I mention his book simply to point out that it is not at all clear that a study of the mind that resembles the study of vision Searle refers to will have much to say about intentionality, as Searle speaks of it. If this is true, it leaves Searle’s claim that intentionality can be explained as a biological feature of brains in serious doubt. Cairns-Smith makes no mention of intentionality. The most obvious place for a discussion of it would be in Cairns-Smith’s description of intelligence; its omission at that point is telling. (Cairns-Smith, A.G. Evolving the Mind: On the Nature of Matter and the Origin of Consciousness. Cambridge UK: Cambridge University Press, 1996) 91. The quote from Victor Serebriakoff is taken from, The Future of Intelligence: Biological and Artificial (Carnforth: Parthenon Publishing, 1987) 9.
out or believe their master is at the door" (Searle 1983, 5). While the fact that a particular conclusion is reached only through philosophical theorizing should not in itself count against the truth of that conclusion, Searle goes on to provide two reasons for believing that it actually is the case that non-language users have intentional states. First, the causal basis of intentionality in such creatures is very much like our own; higher mammals, at least, often have sense organs and brains much like human sense organs and brains. And second, our ability to explain and predict their behavior depends on our willingness to attribute intentional states to them.18

One could hold a position, such as Searle's, that takes intentionality to be a product of organic brain tissue, while supporting a position similar in regard to intentional content to that of Norman Malcolm, discussed above.19 Such a position would claim that while many animals have brains which produce intentional content, only language-users are capable of having thoughts with specifiable content. However, Searle certainly does not take this line and there is nothing in his theory that forces one to adopt it. Searle's basic claims, then, are compatible with the idea that non-language-users have thoughts.

18This is an interesting claim for Searle to make, given that he concludes from his Chinese room thought experiment that what a system seems to be doing from the outside does not indicate whether or not that system has intentional states. As he puts it, "Now the point of the story is simply this: by virtue of implementing a formal computer program from the point of view of an outside observer, you behave exactly as if you understood Chinese, but all the same you don't understand a word of it (Searle 1984, 32-3). There is clearly a tension between this position and his claim that animals have intentional states, because they act in ways we could not explain without attributing such states to them. Surely, it would be the case that if some person were good enough at following the rules for manipulating symbols that he or she behaved exactly as if she understood Chinese, then we would have to attribute intentional states to that person in order to predict his or her behavior. If this does not lead us to actually claim that person in the room understands Chinese, it should not be used to support the conclusion that animals have intentional state. This tension between Searle's claims was pointed out to me by Ronald de Sousa.
19See above, 9-10.
While Searle argues that intentionality is a biological property of brain tissues, functionalists offer a general account of the nature of intentional content that does not depend on any particular type of organic material, but rather on the organization and relations between the parts of certain types of systems. I will use Jerry Fodor's position as an example of a functionalist theory of mind. Fodor expresses an attitude, very different from Searle's, toward the place of intentionality in the physical order:

I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of spin, charm, and charge will perhaps appear on their list. But, aboutness surely won't; intentionality simply doesn't go that deep. It is hard to see, in the face of this consideration, how one can be a Realist about intentionality without also being, to some extent or other, a Reductionist. If the semantic and the intentional are real properties of things, it must be in virtue of their identity with (or maybe supervenience on?) properties that are themselves neither intentional nor semantic.20

Searle might agree with the first part of the above claim — digestion and mitosis would probably not make a physicists' list either. Searle sees intentionality as a natural property, but he is clearly not interested in attempting to show how a complete reduction of the intentional to pure physics might be possible. Biological properties of organisms are real enough for Searle and, for him, intentionality is securely on the level of those properties. Biological naturalists, such as Searle, need to worry about eliminative materialists about as much as biologists need to worry about eliminative physicists (those who would eliminate biology in favor of pure physical descriptions). Of course, if one is not as confident as Searle about the idea that the intentional is

created by, and realized in, the brain, theories of how the intentional might be naturalized are of interest.

The driving idea behind functionalism is that physical states can instantiate intentional or semantic content in virtue of their causal connections to other states, behaviors and stimuli. According to functionalist accounts, to be in a particular mental state is to have a physical symbol (or set of symbols) in one's brain that has the content of that mental state. Fodor claims that, "You connect the causal properties of a symbol with its semantic properties via its syntax" (Fodor 1987, 18). According to such a picture, there are, then, three levels of description of an intentional being. There is an intentional or semantic level at which mental states are described in semantic terms and are related to each other in virtue of their meaning and psychological laws. There is a middle level at which we can describe a state in terms of its syntactic relations to other states. And there is a physical level at which the other levels are physically instantiated and which can be described in purely causal terms.

The traditional model of how this is supposed to work has been the computer. Computers "...just are environments in which the syntax of a symbol determines its causal role in a way that respects its content" (Fodor 1987, 19). In computers, it is clear that we have physical instantiations of syntactic relations that can be described in semantic and intentional terms also (e.g. as playing chess, or displaying something about the meaning or spelling of a word). It is also clear that a set of syntactic relations can be represented in any number of different physical materials and
configurations. A similar relation holds between the level of syntax and the level of intentional description: many different syntactic configurations will result in programs that are indistinguishable from each other by those using the computer.

According to Fodor, the relationship between mental states and brain states is one of supervenience. As Fodor describes supervenience, “States of type $X$ supervene on states of type $Y$ iff there is no difference among $X$ states without a corresponding difference among $Y$ states” (Fodor 1987, 30). Hence, according to the functionalist account, one’s mental state could not change without one’s physical state changing as well. The supervenience of mental states on physical states provides an explanation of mental causation. Each physical instantiation of a mental state is a tokening of that state type. Since physical tokenings have causal powers, they can interact with other such states to bring about new states and behaviors. According to Fodor, these physical instantiations explain mental causation. If mental states are not physical states, Fodor argues, they could not have causal powers: “...if mind/brain supervenience goes, the intelligibility of mental causation goes with it” (Fodor 1987, 42).

Most of these claims are common to all versions of functionalism. In addition to these claims, Fodor, along with some other functionalists, claims that mental representation takes place in a language of thought (LOT). As Fodor puts it, semantically evaluable formulas:

...contain semantically evaluable subformulas as constituents; moreover, they can share the constituents they contain, since, presumably, the subexpression that denotes ‘foot’ in ‘I raise my left foot’ is a token of the same type as the subexpression that denotes

21
"foot" in 'I raise my right foot.' (Similarly, mutatis mutandis, the 'P' that expresses the proposition P in the formula 'P' is a token of the same type as the 'P' that expresses that proposition P in the formula 'P&Q'.) If we wanted to be more precise, we could say that the LOT story amounts to the claims that (1) (some) mental formulas have mental formulas as parts; and (2) the parts are 'transportable': the same part can appear in lots of mental formulas (Fodor 1987, 137).

What it means for mental representations to take place in a language of thought is that those representations have syntactic structure such that complex representations can be broken into constituent parts, and these parts can be combined to form new representations according to syntactic rules.

Fodor argues that when functionalism is considered in conjunction with the LOT hypothesis, it provides explanations of how certain features of thought exist. For example, the productivity of thought (the ability of one thinker to produce any of a limitless variety of thoughts) can be explained if the productivity of thought is understood to be much like the productivity of language. In a language:

...each sentence can be identified with a certain sequence of sub-sentential constituents. Different sentences correspond to different ways of arranging these sub-sentential constituents; new sentences correspond to new ways of arranging them. And the meaning of a sentence—the proposition it expresses—is determined, in a regular way, by its constituent structure.21

That sub-sentential constituents can be combined and recombined in an infinite number of ways, explains how the language is infinitely productive. If thoughts are mental symbols, then they could be combined in various ways to create the possibility of the existence of an infinite number of thoughts as well.

According to Fodor, if one adopts the LOT hypothesis, one can also explain how it is that mental processing takes place. If there are mental symbols, the brain can manipulate these in ways similar to the ways symbols are manipulated in computers. This, Fodor claims, is a step toward explaining how chains of reasoning occur.

It could be argued by one who accepts the LOT hypothesis that non-language-using animals do not have a language of thought for the following reason. The language of thought hypothesis, among other things, explains how language has the characteristics it has. Human language involves expressions of sentences in the LOT combined with a set of public rules of use which vary depending on which spoken language one learns. If this is the case, then we might expect any creature that has a language of thought to be able to use a public language, since it would only involve translating one's sentences into public language from mentalese. Translation from mentalese into natural language would be a significant step, but not nearly as difficult as learning to form sentences in the first instance. Non-human animals never make this move. Thus, we have an indication that they have no language of thought and are not thinking things.

However, things would not look good for the LOT hypothesis if this line of argument turned out to be correct. In cognitive science, the language of thought hypothesis is used to explain not only linguistic capacities, but many other abilities of both animals and humans. In order to explain animal behavior, theorists of Fodor's ilk depend heavily on the idea of mental representation. It is by internally representing the world that animals find their way around, and manipulate objects, to suit their
needs. The LOT hypothesis purports to be a way of explaining what some of the properties of internal representation must be. And, if the LOT hypothesis applies only to the mental processes that directly produce speech, then it is not very helpful in explaining the levels at which much of human decision making must take place. If the LOT hypothesis does not apply to non-language-users, or the non-language-oriented mental processes of humans or other animals, it is difficult to see how it could provide a satisfactory general explanation of how it is that thoughts have content.

Fodor's own position on animal thinking is given as part of his attempt to refute the claim that thoughts occur within natural language. Fodor dismisses the idea that thought takes place purely in public languages in the following way: "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." Arguing in favor of the idea that non-language-users think, Fodor asserts that:

...there are homogeneities between the mental capacities of infraverbal organisms and those of fluent human being which, so far as anybody knows, are inexplicable except on the assumption that infraverbal psychology is relevantly homogeneous with our psychology (Fodor 1975, 57).

As an example, Fodor cites the difficulties humans have in sorting tests which involve disjunctive and negative concepts as opposed conjunctive concepts, and points out that:

Animals, too, typically find (what we take to be) disjunctive concepts hard to master. We can account for this fact if we assume that the

---

representational system that they employ is relevantly like the one that we employ (e.g., that an animal conditioned to respond positive to either-a-triangle-or-a-square represents the reinforcement contingencies disjunctively, just as the experimenter does) (Fodor 1975, 57).

The hypothetical argument against animals having a language of thought that I presented above claimed that non-language-using animals could not have a language of thought, because if they did, they would also speak. Consideration of the above comments by Fodor show that although there may be some reason to doubt that non-human-animals share a language of thought similar to ours, these concerns are overridden by the necessity of explaining behavioral similarities between human and animal behavior.23

I now turn to Dennett’s position, according to which, what it is to have a thought is to be interpretable if considered as an intentional being from the intentional stance. When one adopts the intentional stance towards a system, one employs what Dennett calls the intentional strategy, in order to explain and predict that system’s behavior. Briefly, adopting the intentional strategy toward a thing involves, first assuming it is rational, then figuring out which beliefs it should have, given its perceptual capacities and its situation, and finally figuring out which desires it should have, given the kind of thing it is. With this information, argues Dennett, the behavior of an intentional system is explainable and predictable if the system is understood as attempting to reach its goals, while directed by its beliefs. Addressing the question of what it is to have beliefs, Dennett claims:

23See also, the preface to Fodor’s Psychosemantics (Fodor 1987) for a short, informal, explanation of Fodor’s position on the issue of whether his cat thinks.
...all there is to being a true believer is being an intentional system whose behavior is reliably predictable via the intentional strategy, and hence all there is to really and truly believing that p (for any proposition p) is being an intentional system for which p occurs as a belief in the best (most predictive) interpretation (Dennett 1987, 29).

It is clear that an animal's behavior can often be predicted by attributions of beliefs and desires. Such attributions come naturally for us and are in many ways indispensable if we are trying to train, capture, or manipulate the behavior of an animal. Animals act as if they are intentional creatures, and so, by Dennett's definition, are intentional systems.

One might argue that, cast this way, Dennett's net is too wide. The intentional stance can be adopted toward many different types of systems, such as thermostats, lightning bolts, and clams, to which we would not want to attribute beliefs. (We might claim that the thermostat thinks that it is 65°F in the room and that it wants it to be 70°F; that the lightning bolt wants to strike the highest ground possible; or that the clam believes that by digging deeper into the sand it can escape from a predator. In each case, these attributions may allow us to predict and explain behavior. But, most philosophers see it as a mistake to attribute beliefs to thermostats, lighting bolts, or clams. It is reasonable to ask for a way to distinguish between things that it is helpful, in certain situations, to treat as if they have beliefs, and things that have beliefs in some more objective sense. We may ask, for example, if there are systems about which we would be mistaken if we claimed they have no beliefs, regardless of the utility or lack of utility in treating them as if they did.
Dennett sets up the following example to show the sense in which he sees the intentional stance as picking out something objectively true about certain systems. Suppose, he asks, there were beings of vastly superior intelligence and mechanical ability who came to visit earth. For them, predicting human behavior strictly on the basis of physical laws might be as easy as it is for human engineers to predict the behavior of a thermostat on the basis of physical laws. Even though they could predict our movements with great precision, "...if they did not also see us as intentional systems, they would be missing something perfectly objective: the patterns in human behavior that are describable from the intentional stance, and only from that stance, and that support generalizations and predictions" (Dennett 25). If the superior alien physicists could predict the exact movements of a stockbroker’s fingers on a phone and the precise movements of his vocal cords, but did not also know that what the broker is doing is ordering stock and that there are an indefinite number of ways this could be done, then the aliens would be missing something objectively true and significant about the broker’s behavior (Dennett 26). Most importantly, it is in this sense that some systems are ‘true believers’ and others, while it may be useful to treat them as believers at times, are not.

The standard household thermostat falls into the latter category. An alien scientist who did not realize that it was designed to measure and control temperature would be missing something objectively true about the thermostat. One who did not attribute beliefs to it, on the other hand, could hardly be said to be making a mistake. Suppose however, Dennett asks, a thermostat were endowed with a visual system that
could distinguish shivering occupants or other things that would inform it of the coldness in a particular room. Along with this, it could be given some information about geography and the ability to conclude certain things about the likely weather in its area on that basis. Then, the ability to choose and purchase the fuel it will burn could be added to its repertoire. The more features we add that connect the machine to its environment in various ways “...the richer and more demanding or specific the semantics of the system...” (Dennett 31). The machine, through its complex interaction with its surroundings, would come to represent its environment and in very specific ways. Eventually, argues Dennett, the extended thermostat will have to be considered an intentional being. However, Dennett warns “There is no magic moment in the transition from a simple thermostat to a system that really has an internal representation of the world around it” (Dennett 32).

Animals represent and interact with their environments in extremely complex ways. It is an open question as to which non-human animals, if any, behave in ways that make it the case that there are objective truths about their intentional states. But, the idea that some non-language-users think is certainly not inconsistent with Dennett’s theory.

None of the three positions described rules out the possibility of animals having beliefs. In fact, both Fodor and Searle endorse the idea that non-language-using animals have beliefs. But, even if a theory such as Searle’s or Fodor’s is accepted, it may turn out that while animals are somewhat similar to humans, they
just do not have the kind of brain tissue or the correct neural organization to be thinking things.

Even if it does not make sense, according to Dennett’s theory, to search for the ‘magic moment’ at which an animal becomes a thinking thing, it would be extremely useful to have criteria that could provide guidelines to help us decide what types of behaviors show that a particular organism or system is an intentional being. It is clear that according to a theory such as Dennett’s, behaving like a thinking thing is all that being a thinking thing amounts to. In not as obvious a way, functionalist accounts also depend on thoughts being expressed in behaviors. As stated above, functionalist accounts claim that physical states can instantiate intentional or semantic content by virtue of their causal connections to other states, behaviors and stimuli. This is because of the fact that physical states can be connected to each other by their causal properties in ways which allow them to be tokens of syntactic types. At the same time the relations between intentional or semantic states can be represented at the syntactic level. The systems to which functionalist accounts apply are those in which syntactic relations follow the same pattern as the relations between intentional states and are instantiated in a physical system. The question of whether a functionalist explanation of a system can be given only arises if there is an intentional or semantic level to be explained. Given this, functionalists are in need of behavioral criteria by which to judge which activities indicate that a thing is a thinking thing and what type of behavior is needed to establish that that being has a particular belief.
Searle is in less need of such behavioral criteria. Searle's analogy between intentionality and digestion is useful in making this point clear. I assume that, if a biologist is curious to see whether or not a particular organism digests food, rather than waiting to observe the organism's patterns of ingestion and excretion, he or she could simply look inside the organism to find whether or not the correct mechanisms are in place to digest food. Similarly, if one knew what to look for according to Searle's account, one could look inside a brain, rather than observing behavior, to find whether or not an animal has intentional states. It is difficult to see, however, how it would be possible to come to know what to look for without some set of behavioral criteria with which to correlate neurological findings. If there were no behavioral criteria to determine which beings have intentional states, then neurologists seeking the biological basis for intentionality could never know when they have found what they are looking for. To refer to another of Searle's analogies, just as it would be difficult to study vision if there were no way of telling whether one's subject is having a visual experience, so too, it would be difficult to study the physical basis for intentionality if there were no way of knowing which things in the world are intentional. So, the importance of finding behavioral criteria that will help decide which beings have intentional states is important even in Searle's case.

I have argued that when presented in broad outline, at least, three of the most influential theories of mind are consistent with the claim that non-language-using animals have beliefs. Both Fodor and Searle actually endorse attributing thoughts to non-language-users, and Dennett's general theory provides no reason not to do so.
None of the theories provides an account of the types of behaviors that would distinguish between animals that have beliefs and animals that do not. I will turn, in Chapter Two, to an investigation of the types of behavior that justify the attribution of beliefs to non-language-using animals.
Chapter Two
Teleological Explanations and the Attribution of Beliefs

I argued in Chapter One that commonly accepted theories of mind are consistent with the idea that animals have thoughts. In this chapter, I will consider what types behavior indicate that an animal has beliefs. Jonathan Bennett provides criteria to distinguish when it is appropriate to apply belief/desire descriptions to an animal from when it is not.

In *Linguistic Behavior*, from which I will draw heavily in this presentation, Bennett attempts to explain what it is about certain behaviors that make them linguistic. His stated project is to come up with behavioral criteria that distinguish language use from other activities. He attempts to use his explanation of belief to help explain meaning which, in turn, plays a role in his explanation of language. I will not attempt to explain Bennett’s conception of meaning or to give a complete account of his theory of belief. Instead, I will focus on his position on the types of behavior that indicate that an animal has beliefs. I will, however, provide some general background information concerning Bennett’s position on belief/desire explanations and teleology in order to situate his discussion of behavior.

Bennett claims that theories of belief and desire are part of a more encompassing theory of teleological explanations of behavior. He credits Charles Taylor for the concept of teleological explanations he refers to. I will begin with an

---

explanation of Taylor’s account of teleology, as given in his *The Explanation of Behavior*, then I will turn to an explanation of ways in which Bennett builds on Taylor’s work; from this point I will move to a presentation and discussion of Bennett’s behavioral criteria for the attribution of belief. Finally, I will argue that Bennett’s position on behavioral evidence that an animal has beliefs is consistent with the theories of mind I discussed in Chapter One.

Bennett uses Taylor’s account of teleological explanations as a starting point for his own work on teleological theories. In turn, he sees belief/desire explanation as a type of teleological explanation. Taylor’s work on teleological explanations is important because it addresses the concern that all teleological accounts are seriously flawed because they depend on references to purposes.

According to Taylor, central to the idea that human or animal behavior is purposive or teleological in nature is the claim that:

> the events productive of order in animate beings are to be explained not in terms of other unconnected antecedent conditions, but in terms of the very order they produce. These events are held to occur because of what results from them, or, to put it in a more traditional way, they occur ‘for the sake of’ the state of affairs which follows (Taylor 1964, 5).

A teleological explanation then, will link an event with the end or the purpose for which the event occurs.

If we are in the market for empirical explanations, it is difficult to see how explanations that make reference to an event’s *purpose* can be adequate. This is because causes must precede their effects and purposes succeed the events they are

---

supposed to explain in teleological accounts. In such explanations, it would seem that purposes could have no directly measurable empirical content. Taylor's account escapes this difficulty by showing how teleological laws can be expressed in empirical terms.

Teleological laws, according to Taylor's account, take the form: \((B \rightarrow G) \rightarrow B\), which can be read: if a behavior \(B\) will lead to the goal \(G\) then \(B\) will occur. Whether or not such a law will apply to a system is shown by Taylor to be a function of features of the state of the system and its environment, both of which are empirically identifiable. The relevant features of both the state of the system and the environment are defined in relation to the goal to be realized if the behavior that the law is supposed to explain occurs. Whether \(B\) will bring about \(G\) is an empirical question concerning \(G\), the system and environment. Furthermore, Taylor argues, whether \(B\) will occur if it will lead to \(G\) is an empirical question as well, again depending on the state of the behaving system and the environment in which it is functioning. In essence, the behavior occurs because, given the system's state and situation, the behavior described by the law is a condition for the obtaining of the goal described by the law. If the state of the system, or the state of the environment, vary in such a way that the behavior will not lead to the stated goal, then the law will not apply.

Taylor gives an example of a predator stalking prey to illustrate how these aspects — the state of the system and the state of its environment — are related to the goal to be achieved. In this case, engaging in stalking behavior is a function of the
state of the predator (the predator must be hungry) and of the predator's environment (it must have a fairly good chance of catching its next meal if it stalks). These features can be identified on empirical grounds and can easily be seen to relate to the predator's goal of obtaining food. As Taylor explains, "The condition of an event $B$ occurring, is, then, not a certain state $P$ [an unobservable purpose], but that the state of the system $S$ and the environment $E$ be such that $B$ is required for end $G$, by which the system's purpose is defined" (Taylor 1964, 9-10). Teleological laws, then, will say that $B$ will occur if the system in question is in a particular type of state $S$ and is in an environment in which $B$ will lead to its desired goal $G$. The components, $B, S, E$, and $G$ are, in principle, independently and empirically identifiable. Thus, Taylor's formulation of teleological laws overcomes the difficulty expressed by the charge that teleological accounts must refer to entities that cannot play a role in causal explanations.

If we can describe teleological laws empirically, one might ask whether or not teleological principles can be eliminated completely. Taylor's account shows that teleological laws can be expressed in a way that does not include any non-observable entities as components. It would seem then, that for any such law, the conditions under which $B$ will lead to $G$ could be described *intrinsically*; that is, they could be described without reference to the fact that $G$ is a desired goal of the system being described. In the case of the predator mentioned above, rather than describing its environment as one in which its stalking will lead to its capturing prey, that environment can be described in terms of its features — or key stimuli impinging on
the predator's sense organs, which are provided by the environment being described. 

$E$ then, will stand for the enumeration of the conditions that describe the set of environments equivalent to the set described by $E \rightarrow B$. Another way of putting this would be to say that $E$ picks out precisely those environments in which the behavior $B$ will occur. Since the environments in which $B$ will occur are those environments in which $B$ will lead to $G$, $E$ can replace $B \rightarrow G$. This can be combined with the claim that the combination of $E$ and $B$ will lead to $G$ resulting in the new law $((E \rightarrow B) \land (E \land B) \rightarrow G)$. $((E \rightarrow B) \land (E \land B) \rightarrow G)$ conveys everything that the original teleological law $((B \rightarrow G) \rightarrow B)$ does, but, in the new formulation, there is no necessary reference to a teleological relation. The teleological law is replaceable by a non-teleological law. So it would seem that teleological laws are completely eliminable.

Taylor argues, along the following lines, that this is not the case. As pointed out above, in cases in which a teleological law (of the form $(B \rightarrow G) \rightarrow B$) fits, any instance which meets the condition 'B is required for G' (Taylor refers to this as $T$) will also fulfill some other intrinsic description $E$. But, this is not to say that $B$'s occurring is a function of $E$'s occurring or that $B$ depends on $E$. As Taylor argues,

...it may be that in other circumstances a situation which fulfils the description $E$ is not followed by $B$, the circumstances being precisely those in which the situation does not also fulfill the conditions for the description $T$; whereas all cases of $T$ may be followed by $B$. (Taylor 1964, 13).

Putting the idea in terms of the predator mentioned above: there could be cases in which we had the best possible description of the stimuli produced by the
environment (the description referred to by $E$ which captures the antecedent conditions of the occurrence of $B$), but there were still instances in which all of these stimuli impinge on the receptors of the predator, yet $B$ does not occur. These would be precisely the cases in which $B$ is not required to produce $G$. In such cases, $B$ would depend on $T$ rather than $E$. Taylor explains:

When the antecedent in a teleological type law $T$ is replaced by a non-teleological antecedent $E$, the assumption is that all cases of $T$ which are followed by $B$ are cases of $E$. But whether this is so or not is an empirical matter. We have no guarantee a priori that we shall discover an 'intrinsic' characterization $E$ which will apply to all cases of $T$...and which will never apply to a case which is not followed by $B$ (Taylor 1964, 14).

Taylor’s point is that the assumption that we will be able to, in any case, enumerate the criteria for identification of the class of environments described by a teleological law by listing features other than those within the teleological relation described by the law, is unwarranted without empirical verification. The only way then, to justify the claim that a teleological law can be characterized in a non-teleological way is to actually do so. In cases in which no mechanistic explanation has been produced — and there are innumerable cases of human and animal behaviors for which we have no such explanations — Taylor claims that we cannot assume that there is one. The upshot of this is that there may be teleological generalizations which cannot be accounted for in mechanistic terms and, in any particular case, it is an empirical question as to whether or not a teleological law can be explained in non-teleological terms.
What Taylor is arguing for is a position in which teleological laws are taken to be legitimate because they are describable in empirical terms, but, in which teleological relations cannot be completely eliminated, because it has not been shown that mechanistic explanations can replace teleological ones in all cases.

Ronald de Sousa attacks Taylor's claim. He argues that it would never make sense to accept that a teleological law is not explainable in non-teleological terms:

...we can't prove that there aren't any such cases, but it would never be rational to accept anything as a bona fide instance. For the most we could know at any particular time is that we haven't yet discovered the underlying mechanism. In light of past progress in finding mechanisms to explain apparently "insightful" or teleological behavior, what could justify giving up the search? 26

According to Taylor's position, unless we have given a mechanistic explanation of a particular phenomenon, we should not simply assume that there is such an explanation. De Sousa's point is that while it is possible that there are non-reducible teleological laws, we could never be in a position in which it would make sense to think that we had found one. Just as it is impossible to be certain that we will never come across a teleological law, for any teleological law, it is impossible to show that no mechanistic explanation of it exists. Our success in finding mechanistic explanations for behaviors that were, at some prior point, explainable only through a teleological approach, should lead us to believe that there is a high probability that all such behavior is explainable via mechanistic laws. Supporters of Taylor's position then, need to do more than show that it has not been empirically demonstrated that there are no teleological laws; they need to provide an argument that explains why we

should not expect that such mechanistic explanations will be found. There is no indication that there is such an argument. So, although it may be possible, as Taylor claims, that non-reducible teleological laws exist, it would be unreasonable to think that any actually do. If one accepts, as Taylor does, that true teleological laws are not reducible, this is extremely bad news. However, there are other accounts of teleology—such as Bennett's, to which I now turn—which do not accept Taylor's position on this point.

While Bennett's account of teleology is based on Taylor's, Bennett's claims concerning the relation of teleological accounts to mechanistic accounts differs from Taylor's in that Bennett explicitly rejects what he calls the rivalry thesis. The rivalry thesis is the claim that, if there is a teleological explanation of some behavior, then that behavior cannot, or should not, be explained mechanistically (Bennett 1990, 72). According to Taylor's position in The Explanation of Behavior, in true cases of teleological generalizations there are no mechanistic generalizations that will explain the teleological generalization in question. De Sousa's argument, cited above, shows that it would never be reasonable to think that we had discovered such a law. If true teleological laws are not reducible to mechanistic laws, and there is no reason to accept that there are any non-reducible teleological laws, then, rejecting the rivalry

---

27 In another article, Taylor rejects the claim that all true teleological explanations are not explainable in empirical terms, and says that "There is no necessary incompatibility between our describing and explaining behaviour by purpose in ordinary life or in the context of scientific theories of the teleological–intentional type … on the one hand, and our being able to give a mechanistic neurophysiological account of them on the other." Charles Taylor, "Reply [to Robert Borger]", Explanations in the Behavioral Sciences, eds. R. Borger and F. Cioffi, (Cambridge: Cambridge UP, 1970), 89. (This passage is quoted in Bennett 1990, 72).
thesis is necessary if one is to build a plausible teleological theory which yields teleological laws.

But, if the rivalry thesis is rejected and it is admitted that there are no events explainable and predictable only by teleological explanation, a question immediately arises: if we can always give a mechanistic account of an event, shouldn’t we always opt for this over a teleological account? If any behavior which can be explained teleologically can be explained mechanistically, then it seems like the teleological explanation is not really an explanation on its own, but shorthand for the more cumbersome mechanistic explanation. And, if so, it appears that while we may be justified in using teleological explanations, we should keep in mind that such use is only for the convenience of not having to do the real work of obtaining the same information through the methods of mechanistic explanation.

Bennett argues against this position and maintains that there is more to be gained from teleological explanations than convenience. Bennett’s position is that: “...if a teleological generalization does work for us – giving us classifications, comparisons, contrasts, patterns of prediction that mechanism does not easily provide – then that justifies us in employing it.”

This may seem like a simple instrumentalist answer, which one might accept on the condition that all such explanations be given up once a proper mechanistic explanation is worked out. Contrary to this idea, Bennett argues that the value of such explanations is not due only to the fact that current mechanistic explanations are incomplete. Different types of explanations rely

---

on different groupings of events. If we take the mechanistic explanation of animal movement to be an explanation according to chemical or biological science, this will group events and causes in a very different way than will a teleological explanation. Chemical explanations will group events into classes of similar chemical events, while a class of events according to teleological explanations may include all events that lead a particular type of animal to obtain food. The events classified as similar according to a teleological explanation may have very little in common chemically.

Bennett claims that:

If the grouping is the best for our purposes, say because we are studying how the animal relates to its ecological niche or trying to protect our crops from it, then we need not apologize for the teleological explanation which yields it (Bennett 74).

Even if we have a fairly complete account of the chemistry affecting an animal's brain and bodily state, and can therefore predict its movements with a fair degree of accuracy, this may not provide the information for which we are looking: "...we may have left wide open the question of whether the animal will pursue prey or flee from a predator; and this may be the very question we want answered" (Bennett 1990, 74). Of course this works the other way as well. We can know that an animal will attempt to escape a predator or that a person will pay his or her debts without having to know exactly how the animal or person will move. Bennett's point is that there are significant relationships that are captured by teleological analysis of events, and that cannot be captured even by an accurate and complete mechanistic explanation. So, while Bennett disagrees with Taylor's assertion that no mechanistic account of true cases of teleological explanation can be given, his view is not that the only value of
teleology is convenience. Instead, he argues that teleological explanations give us something that mechanistic accounts cannot; they give us classifications of event types that allow us to answer questions as to whether an animal will seek food, flee a predator, attempt to hide itself, etc. If these are the questions we are trying to answer, rather than questions concerning the precise movements an animal will make, teleological explanations will be far more useful than non-teleological accounts, and this utility is the product of the ability of teleological accounts to capture important aspects of the behavior and activities of certain kinds of beings that non-teleological accounts cannot.

The justification for the use of teleological laws given by Taylor depends on the claim that there is no certain guarantee that there are mechanistic explanations which will replace all teleological ones. As De Sousa’s argument shows, however, there is no reason to expect that such mechanistic laws will not be found with continued investigation. Given that, it might seem that the use of teleological explanations is not justifiable except by reference to their usefulness for the period in which mechanistic explanations are being developed. But, Bennett argues that there is a firmer justification for the use of teleological laws. While it may be that any event which is predictable via teleological laws is also predictable via mechanistic laws, the groupings, classifications, and relations between events provided by teleological schemata may suit our needs and answer questions in which we are interested better than mechanistic frameworks. If we ask whether two predators, a wolf and a lion for example, are engaged in a similar activity such as stalking prey, the teleological
approach may be far more useful than the mechanistic. This is important for Bennett to establish since he takes belief/desire explanations to be a species of teleological explanations and uses his account of teleology as the basis for his account of belief. In order to develop this idea, Bennett adds to Taylor’s account of teleological laws. The concept of registration is perhaps Bennett’s most significant addition to Taylor’s formation of a teleological law. I will describe this addition and the basic structure of teleological laws according to Bennett and then turn to his explanation of belief.

In Bennett’s terminology, if an animal has a goal G and is in a position such that, if it performs a certain action a, it will obtain G, it is said that the animal is a/G. It is Taylor’s view that in order for a teleological explanation to be applicable to a system, it is necessary that for that system, ‘...there is a reliable generalization of the form: ‘For any a, whenever it is a/G it proceeds to do a if that is within its physical competence’’ (Bennett 1983, 356). (This corresponds to “ (B → G) → B” in Taylor’s notation.) Bennett argues however, that there are no laws of this basic form, since an animal couldn’t be expected to act on the fact that it is a/G, even if it had the ability to perform such an action, unless it were informed of the fact that it was in this state.

Bennett gives the following example: if a lab assistant is told to feed a rat only if it walks seventeen inches forward then lies down for seventeen seconds, unless the rat is also informed of this, we would not expect the rat to get fed (Bennett 1990, 46). Bennett argues that this example will refute any law that says that the rat, when hungry, will act to the extent of its competence to do whatever is required to get food. Furthermore, he claims, that examples of this type will be a problem for any law that
simply claims that if an animal is \( a/G \) then it will do \( a \) if it can. So, showing that an animal is \( a/G \) is not enough to show that it will, in fact, do \( a \).

To overcome the difficulty, Bennett adds to his formulation of teleological laws the qualification that the system to which the law applies must register that doing \( a \) will bring about \( G \). The resulting form of teleological law is:

\[
(x)(f)(t)((Rx \& Gx \& x \text{ registers that } fX \& Gx \& fX \to fx \& t + d)
\]

\( R \) refers to the animal being in a state such that it has the ability to perform action \( a \) and has \( G \) as a goal. Bennett explains: "As a mnemonic, think of \( 'Rx' \) as meaning 'x' is ready'; but really \( R \) is a complex predicate with longstanding components ('is a wolf') and well as episodic ones ('is hungry')" (Bennett 1990, 39). What \( R \) does then, is specify the qualities that pick out the animals or things to which the law applies. So, what the law formalized above says is that any thing which, at time \( t \), is \( R \) (has specified qualities which define it as a being that has \( G \) as a goal and the ability to obtain \( G \)) and registers that it is in a state such that, if it does an action \( f \) it will obtain \( G \), will at a later time, engage in action \( f \), which will allow it to obtain \( G \). An account of an animal's capacities for registration will allow us to distinguish between instances in which an animal is \( R \) and \( a/G \), and performs \( a \), and instances in which an animal is \( R \) and \( a/G \) and does not perform \( a \).

Bennett uses "registers" rather than "perceives" in order to capture the fact that "...even quite lowly animals may be modified by some environmental fact which somehow impresses itself upon them," (Bennett 1990, 49) while avoiding questions concerning what lower animals really perceive. For example, a piece of frog behavior
might be explained through reference to the fact that the frog has somehow taken in the fact that it is swimming/eating. If so, Bennett goes on:

...I could say that the frog sees that it is swimming/eating; but many thoughtful people hear this as meaning that the frog thinks, going by what it sees, that it is swimming/eating; and they are naturally reluctant to credit a mere frog with thinking that \( P \), for any \( P \) (Bennett 1990, 49).

To avoid such disputes, he uses a term that implies less; it should apply not only to animals that have actual beliefs, but also to the “quasi-sensory states of grass and self-guiding missiles, if the behavior of these is a fit subject for teleological explanation” (Bennett 1990, 52).

One, of course, may ask on what basis registrations are attributed and how we may know what registrations an animal has. As Bennett acknowledges, there is a difficulty in developing such an account of registration in that we cannot explain what an animal registers unless we observe its behavior and know what its goals are. If we observe behavior, but have no idea why an animal is doing what it is doing, then we can have little idea of which part of the stimuli that an animal is receiving is having an influence on its behavior through the animal’s registering of it.

At the same time, we might ask how we can know what its goals are unless we know what it is registering. As Bennett illustrates, we can’t know that an animal is climbing the tree because it saw a nest unless we know that its goal is to eat eggs or, on the other hand, we can’t know that its goal is to eat eggs unless we know that its action is a result of its seeing the nest (Bennett 1990, 48). As the example indicates, the difficulty applies to the identification of goals as well as the identification of cases.
of registration, since goals cannot be defined by the examination of behavior unless we have some idea of what the animal to which we are attempting to attribute goals is registering. The only way to overcome the problem is to develop theories of registration and goals together.

Of course, in developing such a theory, we need to start somewhere. In actual applications of the theory to the behavior of animals, one would most likely begin with basic ideas and observations and use them to develop hypotheses. However, as Bennett points out, one could even start with an assumption as naïve as, animal a has goals and registrations which are the same as those of a normal human being, then make predictions about how a will behave in given situations. When these predictions are mistaken, either the understanding of a’s goals or a’s ability to register information from the environment might be adjusted. There must, of course, be reason to adjust one of these rather than the other.

The development of the theory could take the following form. Assume that we have a theory, in which we are confident, that tells us that if animal a is R and it registers that it is b/G, then it will do b. Based on the theory, we predict that in a particular type of situation the animal will do b, but when we observe the animal in that situation we find that our prediction is false. In this situation, we might first reconsider our claims about the animal being R. We could observe the animal when it is in different states or conduct experiments to determine if the some other state of the animal R* should replace R. If this works, then the problem is solved. But, if we are confident in our claim that a is R and that we have defined R correctly, we must
decide either that the animal did not register what we thought it would register, or that it did not have the goal we thought it had.

The hypothetical instance discussed above, involving an animal that climbs trees and eats eggs from nests, is a concrete, if artificial, example of how we can discern which of the two aspects of the theory, goal or registration, must be adjusted. Suppose our theory has led us to hypothesize that if animal $a$ is $R$, $a$ will climb trees when $a$ registers that climbing trees will lead to its goal of eating eggs. On such an occasion, however, we observe that $a$ does not climb a tree in which it seems that a nest is clearly visible. The explanation of the mistaken hypothesis could be that the goal that our theory attributes to $a$ is not actually the goal of $a$ — maybe $a$ only eats the eggs of a particular species and the eggs in the nest are the wrong sort of eggs, or maybe $a$ is not interested in eggs at all but in pieces of food of a certain size. On the other hand, the misleading difficulty might be in the attribution of registration abilities to $a$. It could be that while this nest is black, all of the other nests to which we have seen $a$ climb are brown, and $a$ does not have the ability to register black nests.

If we want to test the goal aspect of the law, we could provide $a$ with eggs of different sizes, or of different species, or we could concoct some other experiment in which food is available to see how $a$ reacts to different food options. If, instead, we want to test $a$'s ability to register that it is in fact in a state of climbing/eating eggs, then we could, for example, present $a$ with a black predator, rather than a brown one.

---

29 Bennett admits that the example is "strained and unrealistic," but claims that it still clearly shows how the difficulty concerning the relatedness of registration and goals can be overcome (Bennett 1990, 51).
This would help us to see, in a context other than egg-eating, if the black-brown difference in the nests is behind a’s acting differently than we expected.

Any test of a’s goals will involve assumptions concerning its registrations. The tests of what types of food a will eat depend on a’s ability to register that it is in a state such that if it walks two steps it can eat food. Similarly, our test of a’s ability to register a black predator can only be informative if one of a’s goals is escaping predators. So, we can never test our understanding of a’s goals completely independently of a’s registrations and vice versa. We can however, by the method sketched above, find out about either of them by combining the elements of both in ways that shed light on one or the other. As Bennett explains:

…we can identify a false epistemic or goal element in our joint behavioral theory, because it will carry falsehood with it through all combinations, except perhaps for a tiny number where it happens to compensate exactly for an error in the other part of the theory (Bennett 1990, 51-2).

Bennett likens this to a situation in which one has to test panes of glass for distortions under the restriction that he or she always has to look through two panes at once. Because flawed panes would carry the distortion through various combinations, it would be possible to tell which pane was flawed by looking at it in combination with various other panes. In the same way, by varying different factors, it could be determined whether a problem in a goal/registration theory lay with the goals or the registration components.

Focusing on registration, rather than goals, what has been explained so far shows how we could figure out the features of its environment to which a is sensitive.
This will allow us to explain why, in cases in which the goal aspects of our theory are accurate, an animal which is in fact $a/G$ does not do $a$. In addition, once we have identified the environmental features to which $a$ is sensitive, we may use this information to help us predict how $a$ will behave in future situations in which such features are either present or lacking.

We may still ask, however, what it means to say that an animal registers that $P$. While information concerning the workings of sensory organs is, of course, important in the development of hypotheses concerning our registration and goal theories, Bennett does not provide a list of organs that exhausts the ways in which sensing can occur. On the contrary, he explicitly avoids a description of registration that limits it to specific types of occurrences — biological or otherwise. Instead he says only that ‘registration’ is introduced “…as a theoretical term, standing for whatever-it-is about a given animal which validates predictions of its behavior from facts about its environment” (Bennett 1990, 52). But, if all we can say is that we know that an animal registers that it is $b/G$ when it is $R$ and $b/G$ and acts in a way that leads to goal $G$ and does not register that it is $b/G$ when it is $R$ and $b/G$ and does not do $b$, the concept of registration does little to help us explain why an animal did or did not act in a way which led to its goals. Any explanation based on only this idea of registration could tell us nothing more than that the animal in question doesn’t pursue the goal we attribute to it under certain conditions. If this were the case, the concept of registration would add nothing to the theory.
As Bennett develops the concept further, it becomes clear that including the concept of registration in teleological theories could help provide more complete explanations of behavior. On Bennett's account, "If a is in an environment which is relevantly similar to some environment where P is conspicuously the case, then a registers that P" (Bennett 1990, 53). Relevantly similar and conspicuously each must be defined within a goal and registration theory for a. Such a theory:

...contains hypotheses about which features of its environments a is sensitive to, and those hypotheses enable us, for many values of P, to pick out a certain class of environments which is made up as follows:

(i) a core of environments where P is the case, and where the features because of which P is the case are the ones to which a is sensitive;

and

(ii) a surrounding ring of environments where P is not the case but which do not differ, in any respects to which a is sensitive, from some members of the core (Bennett 1990, 53-4).

There will be environments in which P is the case but in which a does not have the appropriate sensory apparatus to be informed of the features which indicate P. It is the distinction between these environments and ones in which P is the case and a is sensitive to the features in virtue of which P is the case that mark the cases in which P is conspicuously the case. There are also those environments in which P is not true, yet all of the features that a is sensitive to do not differ from those in which P is the case. This is what Bennett means by a relevantly similar environment. Each environment described in (i) is an environment in which P is conspicuously the case. The environments described in (ii) are those that are relevantly similar to some
environment described in (i). Of course, it is also true that members of (i) are relatively similar to themselves.

In response to the worry expressed above about whether the concept of registration added any explanatory power to Bennett’s theory, Bennett can reply as follows. If we could know that an animal had registered a piece of information only if it was $a/G$ and performed action $a$, then registration could not be doing any explanatory work in the theory. But, the above classification of environments, when combined with a goal/registration theory, will lead to predictions that would not be possible without the employment of the concept of registration. For a given value of $P$, we can identify the environments in which $P$ is the case. It is possible then, if we have a reasonable grasp of the kinds of things to which $a$ is sensitive — which we get from working with our goal/registration theory for $a$ — to figure out in which environments $a$ will register that $P$ and behave accordingly, even if $P$ is not the case. This is important because while our claims concerning $a$’s registrations are still based on $a$’s behavior, registration is no longer linked directly to the obtaining of goals, but instead, to the features of $a$ and $a$’s environment that explain $a$’s actions in some cases in which no goal is obtained. Since these predictions could not be made without the employment of the concept of registrations, registration is not eliminable from the theory. We should expect teleological theories to explain behavior in situations in which the goals we attribute to a subject are obtained, as well as in situations in which the goals we attribute are not obtained and situations in which it is clear that the
environment is not as it appears to be to the subject. If a teleological theory is to work in such situations, it will need a concept of registration, or some similar concept.

Justifying the use of teleological explanations and explaining the concept of registration are only steps in Bennett's explanation of what behaviors justify the attribution of beliefs. Not all teleological explanations involve the attribution of belief-like states, and having the ability to register propositions is far from having the ability to have a belief. But, belief/desire explanations, Bennett argues, are a species of teleological explanations, and belief is a species of registration. The most important aspect of Bennett's work, for my project, is his explanation of what distinguishes the cases to which belief/desire explanations apply from the cases in which they do not.

Bennett offers two conditions that serve to differentiate between those teleological systems that have beliefs and those that do not. The first of these is that in order to have a belief that \( P \), an animal must be highly educable with regard to many different kinds of information, including information in the area into which \( P \) falls. The second is that in order to have a belief an animal must be inquisitive. I will discuss what Bennett means by “educable” and “inquisitive,” then provide an explanation of how they fit into Bennett's overall picture of belief.

A system is “educable”, according to Bennett, if it has the ability to change the propositions it registers in particular types of environments: “If for a while \( a \) registers \( P \) whenever it registers that its environment is of kind \( K \), and then later it does not register \( P \) in those circumstances because it has learned better, then \( a \)
manifests educability” (Bennett 1990, 84). If a system can “learn better,” its behavior can be modified on the basis of experience so that its goals will continue to be met even if a changing environment requires changes in its behavior. Bennett provides an example in which an animal sees edible meat dug out of the ground whenever it sees a shovel, then comes to be disposed to dig near all shovels it sees. It may then be said to have been educated into this behavior (Bennett 84). Similarly, an animal might be educated out of a particular behavior, such as digging when it sees prey run into a hole, if such behavior repeatedly fails to lead to the predator actually catching its prey.

When an animal is educated into or out of a certain type of behavior, the causal relationships between its registrations change. In the first example mentioned above, registrations of the presence of shovels come to cause registrations by the animal that it is digging/eating. The second example involves the possibility that an educable animal that registered that it was digging/eating on occasions in which it saw some type of prey run into a hole, could stop registering that it is digging/eating on such occasions.

Educability then, is limited to registrations that are based on other registrations. Bennett provides the example of a predator — perhaps a polar bear — which registers that a certain dark patch is a shadow. Then, on the basis of this, registers that there is a seal on the rock, and, on the basis of this that it is climbing/eating. The predator, if educable, could, with experience, come to stop registering that there is a seal on the rock, in cases in which it registers the existence
of a dark patch. Or, if it turned out that it could not climb fast enough to catch seals before they could escape, it could come to stop registering that it is climbing/eating whenever it registered that there was a seal on a rock. If a registration were not based on any other registrations, it could not be affected by success or failure in the satisfaction of the goals.

I now turn to Bennett’s use of the term “inquisitive.” In Bennett’s terminology, an animal is inquisitive if it is disposed to seek information. For Bennett, the goal of seeking information is *epistemic enrichment*. In Bennett’s terminology, an animal is epistemically enriched if it registers, as open to it, a course of action that it has not previously registered as a possible course of action for itself, and that course of action is better — has a higher probability-and-value measure — than any courses of action the animal had previously registered as available. In such cases, the animal adds something useful to its epistemic store.

It is possible for us to know that, for a particular action $F$, an animal, $a$, is $F/\text{enriched}$ (doing $F$ is sufficient for it to become enriched). For example, $a$ may be standing behind a rock and we may know that if $a$ walks around the rock, $a$ will register information which is related to the location of prey. In addition, such information — that walking around the rock will lead to information relating to the location of prey — may also be available to $a$. So, $a$ may register that it is $F/\text{enriched}$. Attributing such a registration to an animal may be a key to explaining its behavior. If $a$ does walk around the rock, our best explanation of $a$’s behavior may be that $a$ registered that it was $F/\text{enriched}$ and that this led it to do $F$. In such cases, the
goal of the animal’s action is the gaining of information. According to Bennett, a behavior is inquisitive if it is “engaged in with the goal of obtaining epistemic enrichment” (Bennett 1990, 84).

Bennett argues that it may be the case that all animals that register many propositions concerning different areas of information, and are highly educable with regard to those propositions, have beliefs. He refers to this as his strong thesis. As he states it, “...a believes that $P$ is true if $a$ registers that $P$ and $a$ is highly educable with regard to many kinds of propositions which do not exclude any important kind to which $P$ belongs” (Bennett 1990, 87). This amounts to the idea that if an animal $a$ is, in general, educable, and this educability obtains for registrations concerning $P$, as well as the other areas about which the animal registers information, then $a$’s registrations that $P$ are beliefs. On this picture, “belief slides smoothly off into registration” (Bennett 88). For Bennett, the question of whether an animal does or does not have beliefs is not clearly disjunctive, but rather, an animal fits into the category of believers to a greater or lesser extent. Animals can be more, or less, educable about a greater, or smaller, number of propositions. The more highly educable about a greater number of propositions, the better the label of belief fits an animal’s registrations.

Bennett says, in support of his strong thesis, that it captures the important characteristics of belief:

...we have an animal $a$ whose behavior can be teleologically explained only by reference to a certain proposition’s being registered on $a$; and this mode of registration is such that what is registered on $a$ (i) depends on $a$’s present perceptions, (ii) may depend also upon its past
experience, (iii) can be altered by its future experience, and (iv) is manifested in behavior which is usually suitable to a's goals, though can be unsuitable because a has bad luck or because it has not properly adjusted it registrations in light of its past experience. Doesn't that sound like a description of belief? (Bennett 1990, 88-9).

It is true of Bennett's criteria for belief attribution that it captures the ways, relevant to belief, that believers act; that the behavior of any creature that acts in such ways is best explained when beliefs are attributed to it; and that there are no important counterexamples of systems which meet the criteria but clearly do not have beliefs. These are good reasons to accept Bennett's position on when attributions of belief are justified.

Fred Dretske offers an argument from a different perspective that provides further support for the claim that educability is an important feature of systems that have beliefs. Dretske begins with a claim that in purposeful action, thoughts control behavior:

Purposeful action — in contrast to mere behavior — requires thought, but thought alone is not enough. To qualify as purposeful, thought must control behavior. It is not enough to be a thinker and a doer. The thinking must explain the doing (Dretske 19).

The place where this sort of meaningful action first arises, Dretske claims, is in cases of animal learning. He offers an example of a bird learning to avoid Monarch butterflies, and other butterflies that look like Monarchs. In his example, a foraging bird attempts to eat a Monarch that is toxic to the bird because it has been reared on a particular form of milkweed. After eating the poisonous butterfly, the bird vomits.

---

Subsequently, the bird avoids eating butterflies that look like the one that it ate before it got sick. After making this adjustment, the bird sees a Viceroy, a butterfly which looks very much like a Monarch, but is not poisonous. The bird does not eat the Viceroy, but flies away. Dretske comments on the example as follows: 31

If the bug it saw happened to be a poisonous Monarch, we could have said that it recognized the butterfly as one of those nasty tasting bugs and avoided it because it didn’t want to get sick again. But what it saw was not a nasty tasting bug. No recognition took place. There was no knowledge. We need a different word. What is it we (philosophers) call some perceptual state that would be recognition or knowledge if only it were true? Belief! So the bird believes the bug it sees tastes bad. This is what it thinks. What it thinks is controlling the bird’s behavior (Dretske 28).

Dretske’s claim that we would attribute recognition to the bird if it had encountered a poisonous Monarch and acted appropriately is suspect. The assumption that we should attribute recognition to the bird is significant, if mistakes in recognition are automatically taken to be beliefs. As it stands, the argument seems, at least, to be circular; if a mistake in recognition is a false belief, then recognition is true belief. If it is not clear whether belief should be attributed to the bird, then there is no reason to think that recognition, as Dretske uses it, should be attributed either.

In Bennett’s terminology, we could say that the environment in which the bird encountered the Monarch and the environment in which the bird encountered the Viceroy are relevantly similar. The concept of registration is helpful in explaining the

31 I should point out that I am not concerned as to whether Dretske’s example accurately represents the abilities of any particular bird. For observations that it does not, and further comments on Dretske’s article, see sections 3.2 and 4 of Achim Stephan, “Are Animals Capable of Concepts?” Erkenntnis 51.1 (1999): 79-92.
behavior of an animal in environments that are different, but seem similar to an animal. But, the concept of belief is not needed.

Dretske’s point becomes more clear when he turns his focus to the fact that the bird has learned to attribute properties to butterflies that look a certain way, and that what it has learned plays a significant role in directing its behavior. As Dretske describes the situation, the bird has an internal representation $R$, which means that a butterfly of type $M$ is present, and the bird wants to avoid butterflies of this sort. Dretske concludes that:

...the causal story looks like this: an $R$ which means $M$ causes avoidance because it means $M$. A meaningful state is not only causing behavior...its meaning explains why it is causing it. Meaning is thus explanatorily relevant to why the bird is behaving as it is (Dretske 29).

Dretske argues that this situation is different from ones that we see in the cases of machines, plants, and simpler animals that do not learn. For plants, machines, and non-learning animals, what counts is not what has happened to them, but what has happened to, or what was done by, their ancestors or designers. Machines and plants, Dretske argues, have meaningful parts in them, and these parts control the behavior of the systems. The difference between them and the bird, however, is that, in the bird’s case, the representation $R$ controls the bird’s behavior because of its meaning $M$.

Dretske’s point is not that one must be able to learn to make mistakes. He discusses an example of plants being ‘fooled’ by uncommon weather at particular times of the year (Dretske 26). Machines and plants, argues Dretske, have meaningful representational states. But the meaning of those states does not explain behavior in the same way that it does for animals that learn. In the case of an animal that learns,
its states have meaning and the meaning of the states explains its behavior. If the animal had had different experiences — suppose it had only encountered Viceroy — its internal representations would have different meanings and it would behave differently. Because of this, Dretske argues, the meaning of the bird’s representations explain why the bird’s internal events produce its behavior. The key, for Dretske, is that in cases in which learning takes place, the meaning of an animal’s representations is derived from the animal’s experience, and this gives meaning a significantly different role from the one it plays in cases in which it is derived only from the history of a species or the mind of the designer of a machine.

Bennett might not agree that Dretske’s bird believes the Viceroy would not be good to eat. Although the bird registers that the Viceroy is of type M, only if it were highly educable about many different types of propositions could it be said to have a belief, rather than a registration, on Bennett’s view. I do not bring up Dretske’s position to offer a rival account, however, nor am I concerned with the differing uses of ‘purposeful action’ or ‘thought’. Instead, I want to draw attention to Dretske’s point about the importance of educability. Bennett points out that educability depends on an animal being capable of having registrations that are based on other registrations. Dretske provides one explanation of why this might be important; if an animal is educable, the meaning of some of that animal’s representations are located within the sphere of an animal’s goals and experience.

For those who take Bennett’s strong thesis to be too inclusive, he offers a weak thesis that might be more acceptable. In his weak thesis, the condition that an
animal must be inquisitive with regard to many propositions, including those closely
related to \( P \), is added to the strong thesis. Bennett's weak thesis is as follows:

\[
\text{a believes that } P \text{ is true if } a \text{ registers that } P, \text{ and is highly educable}
\]

about many kinds of propositions including ones like \( P \), and \( a \) is

inquisitive with regard to many kinds of propositions which do not

exclude any important kind to which \( P \) belongs. Perhaps \( a \)'s

registering that \( P \) did not result from inquiry, but \( a \) must generally be
disposed to enquire into classes of matters which do not systematically

exclude \( P \) (Bennett 1990, 89).

The introduction of the condition that in order for an animal to have beliefs it must
also be inquisitive makes the second thesis weaker, in the sense that it reduces the
scope of the term "belief" so that it applies to fewer types of animals. The claims
made in support of the strong thesis still hold — namely that belief is based on
present perceptions and past experience, is alterable by future experience, is
manifested in behavior, and is closely tied to goals. However, animals that are not
inquisitive — and thus can be presumed not to have the sophisticated ability to
manipulate, and to some extent control, their own epistemic states — are considered
not to have beliefs under the conditions of the weak thesis.

Bennett does not indicate that there may be difficulties with the strong thesis,
only that the weak thesis will suit his purposes as well as the strong thesis. Because it
extends beliefs to a smaller group of systems than the strong thesis, and the general
trend among philosophers seems to be to limit the extension of belief even further, the
weaker thesis will be less likely to be attacked. Bennett's fundamental concern is to
present an account of linguistic behavior. He attempts to fit linguistic behavior into a
larger category of meaningful behavior, and for this larger project, the strong and the
weak theses do an equally good job; Bennett merely needs a conception of belief to explain meaning, and nothing is lost by limiting the application of belief to inquisitive animals. So it makes sense for Bennett to use it if he thinks it will raise fewer objections than the strong thesis.

My project is much smaller in scope than Bennett’s. I am interested only in defending the claim that it is, in principle, possible to accurately attribute content to the beliefs of some non-language-using animals. I will not attempt to show that squirrels, dogs, or any other species, have beliefs, or to set up criteria to ensure that those creatures are categorized as having beliefs. I am far more interested in seeing how the content of the beliefs of a being that does not use language can be known and described. So, similarly to Bennett’s project, mine is not hindered by choosing his weak thesis rather than his strong thesis as a starting point.

It is important to my project that Bennett’s account is compatible with each of the three of the accounts of mind — Dennett’s, Searle’s, and Fodor’s — discussed in Chapter One. What Bennett tries to do is to find “sufficient conditions in non-linguistic behavior for the attribution of intention (or purpose) and belief” (Bennett 1990, 32). But, once those conditions have been identified, questions concerning the nature of thought can still be asked. The issue at stake in the discussions of Dennett, Fodor, and Searle outlined above concerns the question of what it is to have a belief. Their answers to questions about whether or not non-language-users think are often independent of what each takes the ultimate nature of beliefs to be.
Bennett’s only support of the weak thesis is that it “looks right” (Bennett 1990, 89). What he means by this is that the class of animals that are both inquisitive and highly educable is the class of animals we intuitively pick out as having beliefs. He claims that these are the animals that we would credit with the belief that $P$ if we had taken them to have registered that $P$. And, he adds that even those who claim that belief requires language talk as if such animals have beliefs, unless they are careful not to (Bennett 1990, 89). This does not seem to be much of a defense. However, the view is plausible enough that unless there are objections that cannot be answered, there is no reason to reject it. Of course, there are objections to the idea that animals have thoughts at all.

If one were to argue, in response to Bennett’s work, with the assertion that some non-language-using animals have beliefs, two basic approaches could be taken. One could accept that Bennett’s account captures all of the features of belief that it should, but argue that there are no non-language-using animals that meet Bennett’s criteria. Or, one could argue that there are crucial features of belief that Bennett’s account does not capture. In this case, the argument would be that the account allows too many systems to be called believers; the requirements set by the account, in this case, would be said to be too weak and the behavior that the account claims is sufficient to identify something as a believer does not do so.

Unfortunately, criticisms of the latter type are outside of the ambit of this thesis and will be left unanswered here. I do not attempt to provide a defense of Bennett’s behavioral criteria for belief. I see Bennett’s criteria for the possession of
beliefs as a starting point from which to address these problems. My work here deals with fundamental questions concerning the former type of criticism. I will not deal directly with questions of the former type such as: do lions or vervet monkeys exhibit behaviors that show them to be educable or inquisitive? Instead, I will focus on a specific set of problems concerning whether or not it is possible to accurately attribute content to the beliefs of non-language-users at all. These are difficulties I see as fundamental to concerns about the attribution of beliefs to animals.

If animals do have beliefs, those beliefs are not likely to be accompanied by a sentence of natural language that passes through consciousness, as is sometimes the case with a human belief. This is a problem for any account of the content of the beliefs of non-language-users if it cannot be explained how such beliefs can be accurately represented. Bennett’s account of what it is to have a belief may be correct, but if it is argued that there is a problem with the idea that the content of a non-language-using animal’s beliefs can be represented accurately, his criteria are not as useful as one might hope. The rest of this thesis is an attempt to deal with precisely this problem. In Chapter Three, as a first step in this direction, I will present Robert Stalnaker’s version of a non-linguistic account of propositions.
Chapter Three
Stalnaker on Possible Worlds and Beliefs

In Chapter One, I argued that three influential theories of mind are consistent with the claim that some non-language-using animals have beliefs. In Chapter Two, I presented Jonathan Bennett’s behavioral criteria for distinguishing between animals that think and animals that do not. My presentation, to this point, leaves a serious concern unaddressed. According to Bennett, non-language-using animals register propositions. As language-using humans, however, we express propositions with sentences, and propositions are often understood by philosophers to be linguistic in nature. This leads to two problems that arise when one attempts to attribute beliefs to non-language-users. First, if propositions are understood to be linguistic in nature, it is difficult to see how they could be accessible to non-language-users. Second, if beliefs are attributed to non-language-users, and the only method of conceptualizing their content is in the form of sentences in natural language, the content of those beliefs is likely to be distorted in our descriptions of them. Because of these concerns, any position according to which non-language-users have thoughts will be strengthened by an account of propositional content that does not cast propositions as essentially linguistic entities. In this chapter I will begin to address such accounts by presenting Robert Stalnaker’s possible worlds account of propositions.32

Stalnaker provides an attractive, non-linguistic, theory of propositional content that can be integrated with work such as Bennett’s in a way that will allow us

to attribute thoughts to non-language users. Because Stalnaker not only offers a theory of propositional content, but also applies that theory to an explanation of belief, introducing Stalnaker's work requires addressing a number of different subjects. I will do so in the following order: I will begin with a very quick presentation of the central idea behind the theory, in order to give a general picture of Stalnaker's understanding of propositions. From there, I will take a few steps back and explain in more detail what I mean by calling his account non-linguistic. As well, since Stalnaker adopts a functionalist position in his account of mind, I will explain the extent to which he agrees with Fodor's position as I explained it in Chapter One. Then, I will explain Stalnaker's theory in more detail, covering his naturalized account of intentionality, and his explanation of the attribution of beliefs and desires. Once I have presented an overall picture of Stalnaker's theory, I will argue that Stalnaker's theory is compatible with Bennett's.

A proposition asserts that certain things are a certain way. In doing so, it also implies that there are certain ways that those things are not. If one considers all of the possible ways things might be, then the content of a proposition can be identified by the set of possible ways things might be in which they are as the proposition asserts, and the set of ways in which things are not as the proposition asserts. Stalnaker builds on the idea that propositions divide the set of possible ways things might be in order to explain propositional content, which he describes in terms of sets of possible worlds.
The central claim of Stalnaker's account is that, "A proposition is a function from possible worlds into truth-values." (Stalnaker 1984, 2) He uses 'function' in the formal sense, as a rule that determines a value relative to an argument in a specified domain. Functions can be individuated by comparing the values they determine for particular arguments: "if functions $f$ and $g$ are defined for the same arguments, and have the same values for each argument, then they are the same function" (Stalnaker 1984, 3). Given the claim that a proposition is a function from possible worlds to truth-values, and the claim that functions can be individuated on the basis of the values they determine for arguments to which they apply, propositions can be individuated on the basis of the values — true or false — which the propositions determine for different possible worlds. The content of a proposition then, is equivalent to, and can be explained in terms of, the set of possible worlds in which it is true and the set of possible worlds in which it is false.

Stalnaker understands possible worlds to be, simply, "ways that things might have been." I will discuss Stalnaker's ontological commitments in Chapter Four. Here, I will say only that Stalnaker does not take them to be, as David Lewis does, concrete particulars that have their own independent and self-sufficient existence, but emphasizes that possible worlds are closely tied to rational action and deliberation. This is part of why Stalnaker argues that his account of propositions is appropriate for explanations of intentional states. He maintains that it captures what is essential to rational activities:

33 Stalnaker attributes the original suggestion that propositions be identified with functions from possible worlds to truth-values to Saul Kripke.
Representational mental states should be understood primarily in terms of the role they play in the characterization and explanation of action. What is essential to rational action is that the agent be confronted, or conceive himself as confronted, with a range of alternative possible outcomes of some alternative possible actions (Stalnaker 1984, 4).

Stalnaker sees the ability to distinguish between alternatives to be the defining feature of rational activity. Propositions, then, according to Stalnaker: "...are simply ways of distinguishing between the relevant range of alternative possibilities — ways that are useful for characterizing and expressing an agent's attitudes toward those possibilities" (Stalnaker 1984, 4). Deliberation between alternative possible actions, on the basis of conceptions of how the world is, and preference for some outcome over another, should be described, according to Stalnaker, in terms of considerations of possible worlds.

It is central to the use to which I put Stalnaker's theory that, according to it, linguistic actions have no special status. This is, of course, to say that an account of propositional content is non-linguistic is to say that it explains propositional content in a way that does not depend on references to linguistic content. Meaning is explained then, without reference to language or linguistic practices, and language is taken to arise in a context in which there are previously existing meaningful relations. Any such account which would explain intentionality, then, would have to do so without reference to language. I will use Stalnaker's approach to providing an explanation of the nature of intentionality to demonstrate how his position differs from linguistic accounts.

---

34 I use 'intentionality' here in Brentano's sense of the term. See Chapter One for a fuller explanation of the term.
According to Stalnaker, giving a naturalistic explanation of intentionality is the central task that a philosophical account of mental representation must accomplish. This involves explaining how it is that: "Some things in the world — for example, pictures, names, maps, utterances, certain mental states — represent, or stand for, or are about other things — for example, people, towns, states of affairs" (Stalnaker 1984, 6). He uses the difference between the way his pragmatic account and linguistic accounts attempt to explain intentionality in order to clarify the deep differences between the two types of approaches:

The pragmatic picture suggests that we explain the intentionality of language in terms of the intentionality of mental states, while the linguistic strategy suggests that we explain the intentionality of mental states in terms of, or by analogy with, the intentionality of linguistic expressions. If we opt for the former course, then to avoid circularity we need an explanation in terms that make no reference to language of the representational character of mental states. If we opt for the latter course, we need an explanation in terms that make no reference to mental states of the representations character of linguistic expressions (Stalnaker 1984, 6-7).

The distinction that Stalnaker draws is not between approaches that assert that intentionality is somehow dependent on public language or interaction and those that claim that it is dependent on some sort of inner language or language of thought. He is instead attempting to distinguish his own account from all accounts that understand intentionality, at least in its most fundamental form, to be a property of language. Thus, philosophers with theories of mind as different as Wilfred Sellars and Jerry Fodor have linguistic accounts. Sellars falls into this category because of his position according to which the intentional content of mental states derives from the content of public language, as does Fodor, because of his strategy of explaining the
intentionality of propositional attitudes in terms of the semantic properties of sentences in a language of thought. According to Stalnaker's account, on the other hand, the intentionality of propositions is explainable without reference to the intentionality of language of either type.

In Inquiry, Stalnaker devotes a considerable amount of effort to arguing that linguistic approaches to explaining propositions will not work. I will not explain his arguments to this effect. While I will briefly discuss linguistic theories in order to help show what is unique about Stalnaker's account, I will say relatively little, about linguistic accounts of propositions. Stalnaker may be correct in his negative assessment of the ultimate tenability of linguistic accounts. The project of this chapter is to work toward a plausible account according to which content can be given to the beliefs of non-language-users. The success of this project is not dependent on the failure of linguistic accounts. It could be the case that both a consistent linguistic account and a consistent non-linguistic account of propositional content can be given. This would surely raise issues, but it is not at all obvious that we would then have to choose between the two types of theories. For example, it may be that linguistic accounts are better than non-linguistic accounts at explaining the propositional content of the beliefs of language-using humans. A non-linguistic account might still apply, and be preferable to linguistic accounts, when we attempt to explain the propositional content of the beliefs of non-language-users. Regardless of

---


36 See Stalnaker 1984, Chapter Two for Stalnaker's principal arguments to this effect.
the success of linguistic accounts, as long as Stalnaker's, or some other, non-
linguistic account works, progress will be made in addressing the concern that it is
difficult to attribute propositional content to the beliefs of non-language users without
a non-linguistic account of propositional content.

A workable non-linguistic account of propositional content will not, alone,
show that any non-language users have beliefs, but it will at least provide a way of
approaching the issue. A theory that describes content such that it is separable from
language allows us to ask questions about what beliefs a non-language-using animal
might have, and this is an important step in determining if any non-language-using
animals have beliefs.

In addition to the possible worlds theory of content, it is fundamental to
Stalnaker's position that propositional attitudes are functional states of rational agents
(Stalnaker 1976, 80). Before returning to an explanation focused on the aspects of
Stalnaker's position pertaining more precisely to possible worlds, I will briefly
explain which of the aspects of functionalist theory he accepts. This will be helpful in
explaining the relationship between functionalism and Stalnaker's possible worlds
explanation of content.

A functional state is, according to Stalnaker, "...a state which is defined or
individuated by its role in determining the behavior of the object said to be in the
state" (Stalnaker 1976, 80). A basic application of this type of theory involves
explaining a piece of an agent's behavior by showing that that behavior would satisfy

---

the agent's desires in a world in which his or her beliefs are true. To use Stalnaker's example, one might explain the fact that Sam is turning cartwheels on the front lawn by pointing out that Sam wants to impress Alice and believes that Alice will be impressed by his turning cartwheels on the lawn (Stalnaker 1976, 81).

In Chapter One, using Jerry Fodor's theory as a model, I claimed that functionalism is an approach to explaining intentional systems according to which descriptions of such systems can be given at three different levels — the semantic level, the syntactic level and the physical level. According to this picture, a physical description of a particular neural structure $x$, at a particular time, can be given. It may also be possible to give a syntactic level description of state $x$. This describes not the state's actual physical properties, but the functional role of the state in relation to other states. Describing such relations is essentially the same as describing a computer program — the same program could be run on many different types of machines and in different machines different physical states could play the same functional role. A description at the syntactic level is not necessarily accompanied by any semantic qualities — there are many examples of programs that can be implemented on different hardware but have no unique semantic or intentional level description. If there is a type of intentional state that correlates in the right kind of way with the functional role that state $x$ plays — that is, is a member of the set of intentional states of which the relation of supervenience holds between the intentional states and the physical states of the system described in terms of their syntax — there is also an intentional or semantic level description of $x$. The key idea behind
functionalism, then, as I presented it, is that physical states instantiate semantic or intentional states by virtue of their syntactic properties. The relationship between intentional states and the physical instantiations of those states is a type-token relationship. A particular type of intentional state, such as the belief that Alice is impressed by cartwheels, may, in different systems, or on different occasions within the same system, be instantiated by any number of different physical states. In each of these instances the physical state is a token of the belief that Alice is impressed by cartwheels. Whether a physical state is a token of a type of intentional state, according to functionalism, depends, in each case, on the relation of that state to other states of the system of which it is a part — the role it plays in causing and being caused by other mental states and behavior.38

One of the strengths of functionalist theory is that it contains answers to important questions concerning mental causation. It is difficult to see how, if mental states are not physical, they can cause behavior and other mental states. The functionalist answer to this difficulty is that intentional states supervene on physical states — for any change in a mental state there will be a corresponding change in a physical state. Every mental state, or at least every occurrent mental state, is instantiated by a physical state. Physical states, of course, have causal properties. Thus, mental states, as instantiated by physical states, can cause behavior and other mental states.

38 For a more detailed explanation of functionalist theory, see Chapter One, 19-21.
As explained in Chapter One, many functionalists take a further step and claim that such representations must take place in a language of thought. Support for this claim is centered on the fact that this would explain important features of thought such as its productivity and systematicity. Language is productive in the sense that, given a finite number of symbols and rules governing the combination of those symbols, an infinite number of sentences can be produced. There also seem to be an infinite variety of intentional state types. The existence of a language of thought would explain this infinite variety of intentional states by claiming that thoughts are combinations of symbols in a language of thought and the symbols in that language can be combined in an infinite number of ways, as they can be in other languages.

Linguistic ability is systematic in the sense that understanding and producing some sentences is not possible without the ability to understand and produce others that are systematically related to those sentences. Anyone who understands “Sam likes Alice” can also understand “Alice likes Sam.” This systematicity seems to be the result of the fact that sentences are constructed from parts that can be combined in different ways — if one understands a sentence by understanding the parts and the rules of combination, other combinations of the same parts will also be understandable. Intentional states also seem to be systematic in this way. If there is a language of thought, the systematic qualities of intentional states can be explained in the same way that the systematic qualities of language are explained.39 The language of thought hypothesis also provides the beginnings of a theory of how mental

39 See Fodor 1987, 148-53, for presentations of both of these arguments.
processing takes place — symbols which have semantic content are manipulated in virtue of their syntactic properties much like in a computer.\textsuperscript{40} For reasons I will explain shortly, I merely mention these arguments rather than discuss them in detail.

Stalnaker maintains that functionalism is correct in the claim that if mental states have causal powers, it must be in virtue of their physical properties, but he rejects the claim that internal representations must take place in a language of thought. In reference to the former claim, Stalnaker argues that any creature that has beliefs and desires must have internal representations, and says of his account of propositional attitudes that it:

\dots is not behavioristic in the sense that it identifies mental states with patterns of behavior. In attributing beliefs and desires, we are attributing certain kinds of internal causal properties which have a structure that tends to reflect the world in ways that make it appropriate to call them representations (Stalnaker 1984, 22).

Stalnaker agrees, then, that propositional attitudes are, in some way, physically instantiated.

As to whether or not there must be a language of thought, Stalnaker is explicitly not committed to any theory concerning the form that internal instantiations of intentional states must take. He claims that internal representations could be in the form of sentences of a language of thought, \textquotedblright\dots but that they also could take the form of pictures, maps, charts or graphs, or (most plausibly) a diversity of redundant forms, none of which are very much like any of the forms which our public representations take\textquotedblright (Stalnaker 1984, 22). Stalnaker maintains that his neutrality concerning the

\textsuperscript{40} See Chapter One, 20-21.
form of internal representations is central to his account of intentional states: "The conceptual separation between form and content is, I think, the central feature which distinguishes the conception of thought implicit in the pragmatic picture from the one implicit in the linguistic picture" (Stalnaker 1984, 23). While some representational system underlies mental states, the *form* that that system takes does not have to mirror the *content* of the mental states it instantiates. Stalnaker adamantly claims that his theory of content does not commit him to any theory concerning the form representations must take.

One might reply to Stalnaker that while it is true that content and form are, to some extent, separable, some forms of representational systems seem as if they would be better suited than others to instantiate mental states. By studying the properties of intentional states we should be able to at least say something about what capabilities the internal representational system underlying those states must have. If we know the capabilities of such a system, we should be able to say something about its features. For example, since the productivity and systematicity of thought needs to be explained, any plausible theory of the system that underlies mental representation will have to be one that posits a type of system that could enable systematic and productive thought. Language of thought theorists point out that sentences also have these properties, while not many other things do. So, even if a non-linguistic account of propositions is accepted, the language of thought hypothesis may be appealing to one who is attempting to explain the characteristics of the form that intentional states must take.
While this is true, Stalnaker maintains that because the content of beliefs and desires is separable from the form in which they are represented it is possible that different agents might have the same beliefs, represented in radically different forms. If Stalnaker is correct, then there is little that could be said about the form of representations that would be relevant to Stalnaker’s claims concerning content.

Suppose, for example, that neuroscience somehow showed that, in mammals, certain structures in the brain serve as symbols in a language of thought, just as Fodor describes. Unless it was shown that no other type of representational system could underlie contentful mental states, the discovery would be irrelevant to Stalnaker’s position. If it were shown that only a language of thought has the capacity to be the physical system in which thought is represented, although Stalnaker would, of course, be obligated to revise his position on the separability of form and content, it still would not show that there is necessarily a problem with the kind of account Stalnaker gives of the content of intentional states.

It would not then be, in any sense, devastating to Stalnaker’s account, if the language of thought hypothesis turned out to be true. Stalnaker’s concern is with the content of representations, not with the form that those representations take: “The aim of the possible worlds definition of propositions is to assign the contents of representations just the structure that is motivated by the pragmatic account of the functional role of representations” (Stalnaker 1984, 23). While Stalnaker does not accept claims that the language of thought hypothesis must be true, he does not attempt to argue that it must be false. As explained above, he says that whether it is
true or not, his interest is in the area of the logical structure of the contents of mental states rather than the form which the representations of those contents take. There is no need, then, to discuss arguments for or against the language of thought hypothesis. Whether or not Stalnaker is correct, even in his claim that form and content are separable, his position concerning the content of mental states may still be supportable.

I will not give an account of issues concerning possible systems of internal representation. This is not because I do not see such issues as important. Indeed, I think that there is much that could be learned about the types of beliefs animals have by studying the way that their sensory organs work and how the information taken in by them is processed. Particularly in the development of theories of registrations and goals, as described in Chapter One, information concerning the way that an animal’s brain works could be as useful as information about the type of sensory organs it has in determining which, of a number of goal/registration theories, is the most plausible. Further, information concerning how the brains of animals work might definitively answer the type of questions with which this thesis deals. But, I will follow Stalnaker in claiming that because the form that representations may take is separable from their content, there is work that can be done concerning content, independently of questions concerning the physiological forms of internal representations.

I now turn to an explanation of how Stalnaker brings together functionalism and his possible worlds theory of content. At the semantic or intentional level of

\[\text{[Footnote 41]}\]

Searle’s Chinese room argument is an example of an attempt to show that computers, of which we do know quite a bit about the internal workings and construction, simply do not have the right kind of internal components to produce intentionality. See above 13-14.
description, Stalnaker claims that propositional attitudes are to be individuated by the roles they play in explanations of rational behavior. The most contentious part of his position is the claim that, if this is true, then the objects of attitudes are possible worlds. According to him, rational decision making is intimately tied to the notion of possible worlds. He claims that, “The picture of a rational agent deliberating is a picture of a man who considers various alternative possible futures, knowing that the one to become actual depends in part on his choice of action” (Stalnaker 1976, 81). It is Stalnaker’s position that all rational activities involve the consideration of possible states of affairs and the recognition of the effects that one’s actions may have in determining which of those possible future states of affairs actually will obtain. The function of desire then, is to divide possible future states of affairs into ones to be sought and ones to be avoided, or according to some theories, to determine the degree of effort with which these possible futures will be sought or avoided. Belief, on the other hand, has the function of determining which possible futures are relevant to the agent and which acts or combinations of acts will lead to which possible states of affairs. Thus, Stalnaker says that, “What is essential to rational action is that the agent be confronted, or conceive himself to be confronted, with a range of alternative possible outcomes of some alternative possible actions” (Stalnaker 1984, 4).

The primary objects of attitudes, then, are possible worlds, or possible states of the world, rather than propositions. To want a proposition to be true is to want the world to be a certain way and to understand a proposition is to know the possible worlds in which it is true and false:
Propositions, the picture suggests, are simply ways of distinguishing between the elements of the relevant range of alternative possibilities — ways that are useful for characterizing and expressing an agent's attitude toward those possibilities. To understand a proposition — to know the content of a statement of thought — is to have the capacity to divide the relevant alternatives in the right way (Stalnaker 1984, 4).

Stalnaker continues by describing entertaining a proposition as directing one's attention to certain possibilities, and contrasting those with others; and distinguishing between two propositions as conceiving a possible situation in which one is true and the other false. If Stalnaker's characterization of the relationship between propositional attitudes and rational activities is correct, then all intentional mental states, including, but not limited to, believing, hoping, or desiring, are explainable as attitudes towards possible worlds. Even those such as idle wishes and passive hopes — which are many steps removed from explicit actions — must be explained, on Stalnaker's pragmatic picture, in terms of the connections they have, through the recognition of ways the worlds could be, to rational action (Stalnaker 1984, 4).

I explained above that Stalnaker sees his theory as giving an account of the content of mental states rather than their form. While Stalnaker does not argue against the language of thought hypothesis, he is not convinced by arguments in favor of it. Considering his account of the contents and roles of propositional attitudes helps make it clear why this is the case. The argument that there must be a language of thought is powerful if it is accepted both that the functionalist explanation of the relationship between the semantic and physical aspects of systems is correct and that propositions have a logical structure similar to that of sentences. If one sees propositions as being linguistic in nature — that propositions are sentence-like, in the
sense that they are made up of components which can be combined and recombined according to certain rules — then thoughts, especially when seen as attitudes toward propositions, must stand in relation to each other in the ways in which sentences are related to each other. The question then asked is, “what kind of system could provide tokenings of the mental states related to each other in the same structurally systematic way that propositions, when taken to be sentence-like, are related to each other?” As an answer to this question, the language of thought hypothesis is very appealing.

Stalnaker does not, however, see propositions as having a linguistic structure. He says of propositions, defined in terms of possible worlds, that they:

...are like sentences in some ways: for example, they stand in entailment relations, can be related as contradictories or contraries. But they do not have constituents which correspond to the semantically simple constituents of sentences, and do not have an analogue of grammatical structure (Stalnaker 1984, 23).

Perhaps even more importantly, propositions do not play the same role for him as for those who take them to be the objects of intentional attitudes. In speaking of his pragmatic picture, Stalnaker says, “This pictures suggests that the primary objects of attitudes are not propositions, but the alternative possible outcomes of agents’ actions, or more generally, alternative possible states of the world” (Stalnaker 1984, 4). It is not surprising that Stalnaker is reluctant to accept the language of thought hypothesis. As argued above, one who accepts that thoughts are attitudes toward propositions, and that propositions are sentence-like, will see beliefs, and other intentional states, as structured and related to each other in the same way that sentences are. Adopting this position, then, will lead one to ask, what kind of a physical system must underlie a
system of propositional attitudes related to each other in this way? The answer given to this question then, will often be that it must be a language of thought.

Stalnaker, on the other hand, needs systems that can instantiate intentional states which stand in certain logical relations to each other, but which are not related to each other in the additional ways that sentences are, by virtue of their grammatical structures. It is not at all surprising then, that Stalnaker accepts that mental states must be physically instantiated, but claims that representational systems other than a language of thought could do the job.

Stalnaker’s claim that propositional attitudes need to be individuated by virtue of the roles they play in explanations of rational behavior is compatible with functionalism as I described it. Still, one could accept this claim of Stalnaker’s, but reject the claim that a reduction of intentional states to anything else is possible. It does not follow from the claim that propositional attitudes should be individuated by the roles they play in the explanations of rational behavior, or even the claim that attitudes can only be individuated in that way, that there is some kind of clean reduction from semantic states to physical states. Dennett, for example, accepts that beliefs and desires need to be individuated in terms of the roles they play in theories of behavior, but rejects the claim that intentional states supervene on physical states in such a way that it would be possible to reduce all mental states to states of the brain in the way that functionalists such as Fodor argue that it is.42 Stalnaker’s claim that propositional attitudes need to be individuated in terms of the roles they play in our explanations of behavior is consistent with many positions that reject the claim that

42 For a specific statement of Dennett’s position on this topic, see Dennett 1987, 71.
mental states must supervene on physical states. I claimed in Chapter One that the conclusions that I will draw in this thesis would not be limited to a narrow range of philosophical positions on the nature of thought. The position I am presenting in this chapter, Stalnaker's account of belief in terms of possible worlds, is at least consistent with the general outlines of theories such as Dennett's and Fodor's.

There is one more issue that needs to be addressed before moving on. One objection to the claim that possible worlds theory can be used to represent animal beliefs stems from the quite obvious observation that what is going on when philosophers do work in possible worlds semantics is very different from anything that could be going on in the mind of an animal. Possible worlds semantics is a complex theory that involves elaborate symbolic representation and the very idea of a possible world is a sophisticated philosophical concept. The theory is supposed to explain and represent the content of beliefs, of both human beings and animals. But, if it is taken to be claiming that either animals or the average human have a sophisticated concept of possible worlds, then the theory is surely wrong. It is clear that neither the average person nor the above average non-human animal, has any explicit notion of possible worlds and it is doubtful that even Stalnaker or Lewis approach many everyday decisions through the framework of possible worlds semantics.

It could be argued that while people and animals do not explicitly think in terms of possible worlds, the possible worlds framework is built into the structure of our representational systems, so to speak. In this case, it would be implicitly guiding
all of our rational actions. This may be more plausible than the idea that anyone explicitly thinks in terms of possible worlds, but does not capture the use I am making of the theory or, I think, Stalnaker's understanding of the role of possible worlds semantics in explaining belief.

Again, it must be kept in mind that possible worlds semantics does not capture the form of the beliefs it describes. If what is captured by possible worlds theory is the content of beliefs rather than the form that representations must take, then possible worlds semantics is simply a way of representing information. Stalnaker's possible worlds theory provides a way of describing the space of possibility that is divided by beliefs. Thus, it provides a way of describing the content of beliefs and, importantly, for my project, a way of describing beliefs that are not themselves structured linguistically.

The use that I am making of possible worlds theory does not involve explaining how animal brains work, or what the representations of the beliefs of animals must be like. All I am trying to do is argue that, if Stalnaker is correct in the claim that rational activities essentially involve carving up the logical space of possibilities, and creatures that do not use language have beliefs, then we can accurately represent the content of those beliefs using possible worlds semantics.

If we are going to individuate propositional attitudes on the basis of the role they play in rational activities, it would be nice to have a clear idea of what rational activities are. I will use the criteria outlined in Chapter Two to distinguish rational systems from all others. Before turning to this however, I will describe the
explanation Stalnaker gives of the relationship between beliefs and desires. In giving this explanation, Stalnaker is attempting to provide a naturalized account of intentionality. My primary objective is to explain how it is that beliefs and desires are attributed according to Stalnaker, and because of this, I will not discuss the general success, or problems with Stalnaker’s attempt to explain intentionality. In addition, I will discuss a difficulty that pertains to both Stalnaker’s and Bennett’s positions, and stems from the interdependence of attributions of belief and attributions of desire. Once this step is completed, I will turn to the question of how to tell whether a being is engaged in rational activities.

Early in this chapter I claimed that according to Stalnaker, the central task of philosophical accounts of mental representation is to provide a naturalized explanation of intentionality. Stalnaker attempts to provide such an account by showing that the relationship between rational subjects and propositions is similar to a less problematic relationship between systems — both organic and non-organic — and propositions. He then argues that our explanations of the relationship between simple systems and propositions will also account for the relation between rational beings and propositions. The type of system Stalnaker has in mind is one, “whose behavior is explainable on the assumption that certain of its states are equilibrium states” (Stalnaker 1984, 12). Stalnaker mentions thermostats as examples of such systems — if the air temperature in a room changes, this brings about, by causal means, an internal change in the thermostat that will cause an air conditioner or heater

\[43\] Stalnaker uses ‘intentionality’ here in Brentano’s sense. See Chapter One, 7-8 for a description of the use of ‘intentionality.’
to cool or heat the room. One could say that this mechanism brings it about that the air in the room and the thermostat are in a state of equilibrium. One might also say that the thermostat itself has reached a state of equilibrium when it is neither causing the heater nor air conditioner to heat or cool the room. Stalnaker also mentions that there are far simpler examples, such as closed volumes of gas for which the kinetic energy in different parts will tend to equalize. For such systems, Stalnaker says, we could define the tendency-to-bring-about relation, which holds between such systems and propositions: "It will be true that \( x \textit{tends-to-bring-about} \) that \( P \) if and only if \( P \) is a logical or causal consequence of \( x \) being in its equilibrium state" (Stalnaker 1984, 12). We could say of a thermostat, then, that it tends-to-bring-it-about that the \textit{temperature in the room is 75°F}, when it is set at that temperature. In this relatively unproblematic way, simple systems can be said to be related to propositions.

Stalnaker claims that understanding the tendency-to-bring-about relation will help us understand belief and desire because of the way that rational agents are related to propositions when they have beliefs and desires. He maintains that:

To desire that \( P \) is to be disposed to act in ways that would tend to bring it about that \( P \) in a world in which one's beliefs, whatever they are, were true. To believe that \( P \) is to be disposed to act in ways that would tend to satisfy one's desires, whatever they are, in a world in which \( P \) (together with one's other beliefs) were true (Stalnaker 1984, 15).

---

44 Stalnaker does not distinguish between these possible descriptions and says that "The equilibrium state might be an internal state of the object or system, or it might be a relation between it and its environment. The object might move toward equilibrium either by undergoing an internal change or by causing the environment to change in the relevant way" (Stalnaker 12). Both seem accurate, depending on whether we take the room to be part of the system, or to be part of the environment with which the system interacts. In any case, his main point is not affected by choosing either description.
According to this picture, to desire something is to be disposed to bring it about, given that bringing it about is in one’s power. ‘Belief’ is defined then, in relation to desire; having beliefs is being in a position in which one will act in ways that will lead to the satisfaction of one’s desires, assuming this is in one’s power. We know what a subject believes by observing what the subject does in order to fulfill his or her desires.\textsuperscript{45} If belief and desire are described in such terms, important characteristics of the relationship between rational beings and propositions are explained. Beliefs, desires and mental representation are accounted for in a way that leaves no aspect of intentionality needing to be explained in an account of mental states.

According to Stalnaker, the central difficulty facing this strategy of explaining belief and desire concerns the fact that the intentionality of belief is explained in terms of desire, while the intentionality of desire is explained in terms of belief. The following question arises, “Is the theory simply a shell game that hides the problem of the intentionality under belief while it explains desire, and under desire while it explains belief?” (Stalnaker 1984, 15). If it is a shell game of this sort, then we may ask if there is a fact of the matter at all concerning which beliefs and desires one may have. Any time we explain a piece of behavior by attributing beliefs and desires to an agent, the hypothesis about an agent’s beliefs can be changed, as long as compensating changes are made in the desires that are attributed to the agent; conversely, any hypothesis about an agent’s desires can be changed, as long as

\textsuperscript{45} This account is oversimplified, of course, and does not address issues such as how to account for the behavior of subjects that have competing desires. In a complete theory, these aspects would be worked out in detail. Laying the foundation for such a theory by explaining the problematic fundamental ideas on which it is based is the project that Stalnaker is attempting. My goal is only to show that the foundation laid by Stalnaker is consistent with Bennett’s.
compensating changes are made in the beliefs that are attributed to the agent. For example, after observing Sam doing cartwheels on the lawn, we hypothesized that Sam wants to impress Alice and that Sam believes that Alice is impressed by cartwheels. But we could also hypothesize that Sam wants to repulse Alice if we adjust our attribution of belief to claim that he believes his showing off his ability to do cartwheels will repulse Alice. It may seem then that according to this theory there is absolutely no fact of the matter about which beliefs agents really have.

Stalnaker responds in two ways. He first claims that this type of argument is not as much of a problem for his account as it is for accounts, such as logical behaviorism, which attempt to reduce unobservable inner mental states to patterns of behavior. The problem is serious for such accounts because if the only factor that determines whether or not a set of beliefs or desires is accurately attributable to an agent is a piece of behavior, then there is no fact of the matter concerning what agents believe, since any piece of behavior is consistent with various sets of beliefs and desires. Stalnaker claims that this difficulty shows logical behaviorism to be false.

He maintains, however, that the same is not true of his own theory. He argues that:

Belief and desire are problematic, not because they are inner states that are not directly observable, but because they are intentional; the analysis is an attempt to explain the intentional in terms of the nonintentional, not an attempt to explain the unobservable in terms of the observable. So we can, without undercutting the job that the analysis is attempting to do, understand dispositional properties such as belief and desire as real causal properties of persons, and not simply as patterns of actual and possible behavior (Stalnaker 1984, 16).
Because the theory takes belief and desire to be actual causal states, even if it is true that different mental states will yield the same behavior, it does not follow that there is no fact of the matter about which states caused that behavior. Stalnaker’s strategy here is to separate the epistemological question of how we can know that a person has certain beliefs and desires from questions about whether there really is an answer to questions concerning what motivates a person’s actions. It is possible then that by assuming that belief and desire are actual causal states, one might be able, as Stalnaker is attempting, to explain the intentionality of belief and desire while not being able to say precisely which desires and beliefs are indicated by which types of behavior.

Stalnaker explains that his pragmatic analysis takes rational agents to be analogous to machines whose inner states are not accessible to us. For such a machine, we could formulate hypotheses about how the machine must work, based on the machine’s observable behavior. It might not be possible to tell, merely by such a machine’s behavior, which of many competing hypotheses was correct. This would not, he argues, show that there was no fact of the matter about which hypothesis was correct. Stalnaker says of his theory that:

In ascribing beliefs and desires to a person, we not only make conditional predictions about how the person will behave; we also commit ourselves to claims about the kind of mechanisms which explain why a person behaves the way he does. The mutual dependence of belief, desire and action is a reflection of the fact that the hypothesized explanation says more than the conditional predictions which it entails (Stalnaker 1984, 17).
While Stalnaker's theory is not committed to a claim about what the forms of internal representations are like, Stalnaker claims that attributions of intentional states commit one to claims about the types of connections one's internal states must have to each other. While the evidence supporting the theory is found in behavior — and is corroborated by the successful prediction of further behavior — the claims that the theory is making about how belief, desire, and action are related are not simply restatements of the conditionals which link patterns of behavior to each other and stimuli from the environment. Thus, in the solution he offers, Stalnaker does not claim that it will always be possible to discover exactly what beliefs a person holds. Instead, he argues that the fact that we may not be able to do this is a result of our lack of access to some relevant pieces of information and should not lead one to be skeptical about there being a fact of the matter about what beliefs a person has.

In Chapter Two, I discussed Bennett's attempt to deal with a problem similar to the one Stalnaker points out and responds to as described above.\(^ \text{46} \) The problem surfaces for Bennett in his discussion of goals and registration. As he explained, it is difficult to see how an animal's goals could be identified unless we know its registrations, and vice versa. The difficulty arises for Stalnaker in his account of the intentionality of beliefs and desires and for Bennett in his account of registrations and goals. Still, it is easy to see that the underlying problem is similar and that the solutions offered by both Bennett and Stalnaker are relevant to the problem as it is presented by both.

\(^{46}\) See Chapter Two, 45-49.
Belief is a species of registration, according to Bennett. So, this problem, since it pertains generally to the concept of registration, is also a problem with belief. But Bennett does not link belief to desire as does Stalnaker. Rather than defining “belief” in relation to desire, which would straightforwardly correspond to Stalnaker’s theory, Bennett explains “belief” as it is related to goals.

In cases in which animals have beliefs, “goal” as used by Bennett and “desire” as used by Stalnaker, are nearly interchangeable.\(^{47}\) The difficulty concerning the intimate connection between goals and registrations discussed by Bennett stems from the same problems that cause the difficulty Stalnaker sees with the relationship between belief and desire. It seems fair to say that if Bennett is correct in his claim that belief and intentionality are species of registration and goal oriented behavior, then the difficulty, as Stalnaker describes it, is a special case of the problem Bennett describes.

Bennett’s solution to the problem is that we begin by forming hypotheses about the registrations and goals of an agent, then test those hypotheses by observing the agent in various situations to see if the set of hypotheses about the agent’s goals and registrations will consistently predict and explain behavior when combined in different ways. For example, in order to test a hypothesis about a particular goal we have attributed to an agent, we can observe the agent in various situations in which

---

\(^{47}\) In sentences such as “It is Sam’s desire that Alice will admire him,” we could replace “desire” with “goal.” In many contexts, this would not change the meaning of the sentence, although it could be argued that in certain cases one could desire something without having it as a goal. In addition, a few grammatical differences make it clear that the two terms are not completely interchangeable. While we can say “Sam desires Alice’s admiration,” we cannot with grammatical correctness say “Sam goals Alice’s admiration.” These differences should alert us to the fact that we must be careful and not assume that the two terms are always interchangeable. Still, the differences in the grammatical uses of the two terms have little to do with the point I am making here.
the goal should be relevant to explaining the agent’s behavior. We may at times manipulate the environment in order to provide the conditions under which we wish to observe the agent as well. The contexts in which the agent is observed acting need to be such that various different beliefs can combine with the goal in the different situations.

After observing Sam doing cartwheels on the lawn in front of Alice, we attributed to him the belief that Alice is impressed by cartwheels and the goal of winning Alice’s admiration. If we want to test our hypothesis that Sam’s goal is to have Alice admire him, we can observe Sam in various situations, or set up such situations and observe him in them, which allow that goal to be combined with different beliefs. We may, for example, observe Alice telling Sam that what impresses her most in a boy is the ability to construct derivations in formal logic. We may then leave a logic text in Sam’s room. If Sam appears to see the logic test, but passes over it and picks up his geography text instead, we have some evidence that, all else being equal, counts against our original hypothesis about Sam’s goal. If, instead, Sam picks up the text and spends hours a day working on his derivation skills, then invites Alice over to do a bit of logic, we have more evidence to support the claim that he is attempting to get Alice to admire him. Of course, this is oversimplified immensely and sorting out the intricacies of complex combinations of beliefs, goals, and behavior is rarely simple, but the fundamentals of the method should be clear.
It may be true that if we look at an isolated piece of behavior and a plausible attribution of beliefs and goals, it may be possible to vary either the attribution of goals, as long as the attribution of beliefs is adjusted accordingly, or adjust the attribution of beliefs, as long as the attribution of goals is adjusted accordingly. But, the situation changes if the agent is observed in various contexts. In a sense, observing an agent in various situations is somewhat like subjecting our attributions of belief and desire to a variety of different tests to confirm, or disconfirm, the attribution. Because different contexts allow one to cross-check one’s attributions, mistakes become apparent.

Still, one might reply that this does not, in fact, solve the problem. For any belief/goal combination, it is logically possible that any change in a goal we attribute can be compensated for by a change in the beliefs we attribute. It might seem, then, that we could do so for any new situation in which we put any subject in question. In turn, we could then adjust other goal attributions to fit with the new belief attributions, and so on. This would mean that for any goal or belief attributed, some consistent set of beliefs and goals would fit with it and any observed behavior. If this is so, no matter how many contexts we observe a subject in, we will be unable, on the basis of consistency alone, to answer questions concerning which beliefs and goals an agent really has. Thus, consistency cannot be the only criterion by which belief attributions may be judged.

In practice, it is not. When we attribute a belief to a subject, we generally have some idea of how the subject came to have that particular belief. Assume we change
our goal attribution to Sam and say that he does not want Alice’s admiration, but, instead, that he wants to repulse her. We could then change our explanation of why he is doing cartwheels to be that he believes that cartwheels will repulse her. But, we need an explanation of why he would believe that. Further, we might need to explain why he spends hours doing logic derivations after he heard Alice say the she is impressed by boys who are good at formal logic. We might be able to come up with explanations, but if the explanations in support of the claim that he wants to impress Alice are more convincing than the ones that support the claim that he wants to repulse her, we have reason to adopt the former claim, even if that latter is also part of a consistent set. Some explanations are more convincing than others because they fit better with what we understand about the kind of information agents register from their environment under particular conditions. When this is taken into account, we can see that the difficulty concerning there being no correct answers to questions concerning which set of beliefs and desires can be attributed to subjects, is not as bad as it may have seemed.48 While it is certainly possible that there will be cases in

48 In response to similar claims, Stalnaker answers in the following way: “Our mental states represent what they represent not only because of the behavior they tend to cause, but also because of the events and states that tend to cause them” (Stalnaker 1984, 18). Stalnaker’s explanation of how this is supposed to work follows from his explanation of what he calls the indication relation between systems and propositions. Some intrinsic states of systems tend to correlate with certain features of the environment. For example, the length of the column of mercury in a thermometer will, if all goes right, vary with the temperature of the air surrounding it in a systematic way. The length of the column of mercury may then be said to indicate the temperature. For any type of system which might indicate some proposition, three things would have to be specified, “first, the relevant set of alternative states of the object doing the representing; second, a one-one function taking these states into the corresponding states of the world; third, the normal or optimal conditions” (Stalnaker 1984, 12).

The difficulty with this type of explanation has been that optimal conditions have been exceedingly problematic to define in a non-circular way. I will not attempt to go into the details of that debate because the problems it poses threaten not only possible worlds accounts or non-linguistic accounts, but all accounts which try to naturalize meaning through an examination of the covariance of inner states with environmental features.
which there is no way to decide between competing belief/goal attributions, in cases in which we have the continued ability to observe a subject in various contexts, it is not likely that problems will arise. This will not be true of cases in which our opportunities to observe subjects are very limited. But, in such cases, the problem is simply that we lack evidence and this does not show that the theory to which we are subscribing is flawed.

So far, I have described the functionalist aspects of Stalnaker's theory, his description of content given in terms of possible worlds, and his explanation of intentionality. Together, these provide at least an outline of a theory of mind. If Stalnaker has produced a workable non-linguistic theory, then he has produced exactly the kind of theory needed by Bennett to fill in some of the details in the picture he paints of meaningful non-linguistic behavior. What I will do in the rest of this chapter is to explain how it is that Bennett's position concerning the types of behavior that indicate that an animal has beliefs and Stalnaker's account of belief, are not only consistent, but complement each other. Stalnaker needs to explain which activities are rational activities; Bennett's criteria will do this and will fit well with Stalnaker's explanation of belief and desire to form a more complete picture.

Bennett's goal is to explain pre-linguistic content. He needs a way of representing sets of beliefs that is not tied to a language; Stalnaker's theory provides that.49

For Stalnaker, intentional states are attitudes toward possible worlds. What it is to be a rational agent is to be able to make choices about how to act based on which

---

49 See Bennett's introduction to *Linguistic Behavior*, xxii, for his comments on Stalnaker.
possible actions will be likely to bring about the most desirable possible states of affairs. Stalnaker maintains that we attribute beliefs on the basis of the roles they play in the explanations we give of rational behavior. The idea here is that rational activity consists in deciding between actions on the basis of desires and recognition of what the results of those actions will be. I take it then, that any being capable of this — deciding how to act by considering one’s goals and choosing the action which one takes to be the best way to reach those goals — is capable of having beliefs.

Stalnaker thinks that non-language-users are capable of having such attitudes, but questions remain as to what signs there are that a non-language-user is actually deliberating in such a way. This is where Bennett’s work comes in. When information from a creature’s environment impacts the behavior of the creature and the creature’s behavior is explainable by reference to teleological laws, Bennett claims that the creature has goals and registrations. I claimed in Chapter Two (pp.59-61) that I would accept what Bennett refers to as his weak thesis about belief: “a believes that P is true if a registers that P, and a is highly educable about many kinds of propositions including ones like P, and a is inquisitive with regard to many kind of propositions which do not exclude and important kind to which P belongs” (Bennett 1990, 89). The general claim is that educability and inquisitiveness mark the distinction between the ability to have beliefs and the ability to register information from one’s environment. If an animal is educable and inquisitive concerning a certain

50 I will discuss Stalnaker’s position on whether or not animals think in Chapter Six.
set of propositions, then, rather than saying the animal registers information, we say that the animal has beliefs about those propositions.

Bennett's work fits well with Stalnaker's. What possible worlds theory, as presented by Stalnaker, does, is provide a way to represent the structure of relationships between mental states. As Bennett points out, very simple organisms and systems such as grass and self-guided missiles, to which we are not often tempted to ascribe beliefs and desires, register information from the environments that they inhabit. Still, we would not likely want to say of such organisms and systems that they are rationally deliberating — deciding how to act on the basis of which action will bring about the most attractive possible state of affairs. If, instead, we consider creatures that, in addition to having goals and registrations, are also educable with regard to those registrations, and inquisitive, then we are looking at a set of creatures that seem to be proper candidates for the attribution of beliefs. As well, it is at least not obvious that such creatures are not deliberating between actions based on the possible states of affairs to which those actions are perceived to lead.

Bennett's main defense for using educability and inquisitiveness to draw the line between registration and belief is that it seems to be in accord with common intuitions about which animals have beliefs. As he puts it, "it looks right," and he points out that even those who believe that language is necessary for belief find it difficult not to talk as if languageless animals that are educable and inquisitive have beliefs (Bennett 1990, 89).
According to Stalnaker’s account, if an animal has beliefs, it has the ability to compare different sets of possible worlds and recognize connections between its actions and the actualization of possibilities. We may ask then, whether or not Bennett’s criteria match up with our intuitions concerning which animals have the ability to represent possible states of affairs and act in ways that seem likely to bring about those states of affairs that are the most desirable.

An animal is educable if it can vary behavior that is not successful in order to act in ways that more often, or more efficiently, lead to its achieving its goals. In possible worlds terms we could simply say that the animal varies its behavior in ways that result more often, or more efficiently, in bringing about the actualization of possible states of affairs that are desirable to it.

If we say of an animal that it is inquisitive, we are saying that it seeks information about its environment. Bennett explains this in terms of epistemic enrichment. An animal is epistemically enriched if it finds a, previously unknown, course of action open to it that will either lead to more satisfaction or will more probably lead to satisfaction than the courses of action of which the animal was previously aware. An animal may register that a particular course of action may lead to epistemic enrichment. Finally, attributing to an animal the registration that a particular course of action may lead to epistemic enrichment might explain an animal’s behavior. For example, under some conditions, the best explanation of an animal’s behavior might say of that animal, after it has been observed walking around a large rock, that it did so to find whether or not a predator was behind the rock. In
such cases, the best explanation of an animal’s behavior claims that the animal has epistemic enrichment as its goal. If this is the case, then the animal is inquisitive.

Bennett’s conception of inquisitiveness can be described in Stalnaker’s possible worlds theory in the following way. We can say that an animal is epistemically enriched if it becomes aware of the fact that a course of action will lead to a more attractive possible state of affairs, or that it is more likely to lead to a desired state of affairs than the courses of action previously known to the agent. An animal, then, is inquisitive if it seeks information concerning the connection between actions available to it and possible worlds or states of affairs those actions will bring about.

We can, then, apply Bennett’s claims concerning how to determine which animals have beliefs to Stalnaker’s possible worlds account of belief. The question remains, however, whether or not Bennett’s criteria, applied in this way, pick out animals that choose between actions based on the possible states of affairs they will help bring about. I argue that if a being is educable and inquisitive about many propositions, including some which are related, then we have evidence in favor of the claim that that being has representations of possible states of affairs.

According to Bennett, if an animal can have a belief that \( P \), it must be educable not only about \( P \), but also about many other propositions, some like \( P \). Organisms may register a lot of interrelated information from their environments, but this, in itself, gives us little reason to understand them as conceiving of possible worlds. There is nothing inconsistent about a system that registers many pieces of
information from its environment, but does not behave as if it has internal representations that could plausibly be interpreted as a registration that those pieces of information are interrelated and fit together to form a possible state of affairs. When we look at any animal that is highly educable about a number of propositions, we might see evidence that, in fact, it has a conception of a world in which a belief may fit. However, it is possible that an animal could be highly educable about a number of related propositions while never having any ability to relate any of that information to the other bits of information it is educable about.

One last time, consider Sam. Through his attempts to impress Alice, or possibly his attempts to repulse her, Sam discovers that he has a special talent for cartwheels, and decides to move to New York and pursue a career as an acrobat. After arriving in the city, he learns how to get from his apartment to a coffee shop, grocery store, library, school, bookstore, and bar — all very close to each other — and gets better at finding the way to each of them, but is completely ignorant of the relationships between the locations of the places. It would likely be apparent to an observer who spent a bit of time with Sam that Sam behaves strangely. It might for example be that Sam takes one route to the bar and another, longer, route to the bookstore, even though the bookstore is next to the bar. It might not be obvious that Sam has no conception of the relationships between each of the places he visits, but this is not likely. It would certainly catch a thoughtful observer's attention if Sam bought a book and went home, only to turn around and walk back to the coffee shop near the bookstore to read it. On the other hand, if, in this situation, Sam left the
bookstore and walked directly to the coffee shop, we would, assuming the coffee shop is not between the bookstore and home, have evidence that Sam is, in fact, able to conceive of each of the locations as being related to the others. Of course, the best evidence we could have of this would be the case in which we show Sam a better way home from the coffee shop and he immediately begins to use this as his new route to the bookstore as well.

The point is not that all possible worlds are complex, and that in order to represent a possible world, one must be able to represent very complex and interesting facts. This could not be the case since the set of possible worlds contains a possible world in which nothing exists and an infinite number of very simple worlds with one, or very few, objects. For this reason, if an animal were not educable about many different, but related, propositions, it would be hard to show that it had any beliefs at all, since one of the important parts of having a belief is being able to relate that belief to other pieces of information.

The concern is rather that, unless a being can represent complex states of affairs, we have no evidence that it can properly be said to choose how to act on the basis of which action is likely to lead to the actualization of the most desirable possible state of affairs. Having the ability to register connections between different aspects of possible worlds and between actions and results seems to be a necessary condition of rational action.

Behavior will often reveal whether or not an animal has registered such connections. If a subject is highly educable about a large number of related
propositions, does not act in inconsistent ways that reveal a lack of ability to connect the implications of the propositions she has learned, and does act in ways that indicate that such implications are recognized, we have evidence that the subject we are observing fits new pieces of information into a conception or representation of a possible world, or state of affairs. Evidence that animals do or do not represent possible worlds, is analogous in many ways to Sam's case as it is described above. The more related propositions about which an animal is educable and the more the animal reflects the effect of learning one proposition on situations involving other propositions, the more evidence we have that the animal has beliefs. The key question to ask is whether or not an animal seems to be integrating pieces of information into a larger system.

If an animal is inquisitive as well, the best explanation of its behavior attributes to it the registration that a particular activity will enable it to become epistemically enriched and the goal of being epistemically enriched. Such an attribution implies that gaining information is a goal of the animal. If an animal has the goal of obtaining information, we have to attribute to it enough of a conception of possible states of affairs for it to register that more information will be useful.

There is no reason to think that there is a sharp cutoff point between animals that have a conception of possible worlds and animals that do not. As I have described things, it is more reasonable to think that there is a continuum, on one end of which, are animals with extremely elaborate conceptions of possible states of affairs, and on the other end of which are animals that do not register much, if any,
information from the environment surrounding them. If it is agreed that thinking creatures choose how to act on the basis of which actions will lead to some possible world becoming actual, we can still ask what evidence could lead us to the conclusion that some animal is capable of such a process. If an animal is highly educable and inquisitive regarding many propositions, some related, then it is difficult to imagine it not having the ability to act because it registers that its actions will bring about certain states of affairs. According to the claims Stalnaker makes concerning rational deliberation, such an animal can be said to be engaging in rational activities.

The stated goal of this chapter was to work toward a viable non-linguistic account of mental content. Stalnaker’s work certainly is a step in the right direction. He provides a non-linguistic account of meaning that can be consistently integrated with Bennett’s work on the subject of distinguishing between animals that have beliefs and animals that do not. I have provided an account of the fundamental ideas in Stalnaker’s account of belief, desire, and intentionality. Last, I argued that Bennett’s criteria for distinguishing animals that think from animals that do not, can work in conjunction with Stalnaker’s possible worlds account of belief. It would seem then, that we need look no further than Stalnaker’s theory for a viable non-linguistic theory of content. I will argue in Chapter Four that this is not the case. Stalnaker’s theory is a step in the right direction; however, I will argue that either Stalnaker’s theory needs to be modified, or that a different theory needs to be adopted to explain non-linguistic propositional content.
Chapter Four
Problems with Possible Worlds

In Chapter Three, I outlined Robert Stalnaker’s possible worlds-based explanation of mental content. In this chapter, I will address objections to Stalnaker’s theory. The most obvious problems with Stalnaker’s account have to do with the fact that possible worlds are generally considered to be metaphysically suspect. In giving possible worlds a central role in his theory of propositions, Stalnaker adopts a position that requires balancing between two types of difficulties. If the account depends upon the actual existence of possible worlds, then it is false, since possible worlds do not really exist. On the other hand, if it uses possible worlds, and merely takes them to be useful fictions, then it doesn’t really provide an account of the nature of propositions. The first task I will turn to, then, is to establish the sense in which, according to Stalnaker, possible worlds exist. I accomplish this primarily through a comparison of Stalnaker’s position to that of David Lewis.

I have chosen to do this for two reasons. The first is that much of Stalnaker’s work on this topic is presented in response to Lewis’ work. The second is that Lewis provides a clearly articulated version of realism toward possible worlds, and comparing Stalnaker’s position to it helps clarify exactly what Stalnaker’s position is. Stalnaker calls his position toward possible worlds moderate realism. After my discussion of what it means to be a moderate realist, I will argue that Stalnaker is realist enough to meet his theoretical needs and moderate enough to avoid problematic metaphysical commitments.
However, Stalnaker’s theory has an additional difficulty that arises from his understanding of the nature of possible worlds. According to Stalnaker, the content of propositions is explained in terms of the sets of possible worlds in which they are true and false. Thus, logically equivalent propositions have the same content since they are true in the same possible worlds. According to the theory, logically true statements like ‘2+2=4’ and ‘there are an infinite number of primes’ have the same content. The theory has a further implication that any person who has one logically true belief believes all logically true statements. While Stalnaker has attempted to address this counterintuitive implication of his theory, I will argue that revisions to his theory and his conception of possible worlds are necessary to provide an explanation of human beliefs. In Chapter Five, I will argue further that this problem with Stalnaker’s theory forces us to look to a different non-linguistic account of content.

Stalnaker uses the realist conception of possible worlds advanced by David Lewis as a starting point and a backdrop against which he can clarify his own position. Lewis employs the concept of possible world in his analysis of counterfactual conditionals, defining the logical operators he introduces in terms of “...how the truth value at a given possible world of a counterfactual conditional is to depend on the truth values at various possible worlds of its antecedent and consequent.” (Lewis 1973, 1) Possible worlds play a central role in Lewis’ analysis of counterfactual conditionals, as they do in Stalnaker’s analysis of propositions. Thus, Lewis, like Stalnaker, is in a position in which he must defend his use of the

seemingly problematic notion of possible worlds. Lewis addresses this by arguing that the existence of possible worlds is not as problematic as it often seems:

It is uncontroversially true that things might have been otherwise than they are. I believe, and so do you, that things could have been different in countless ways. But what does this mean? Ordinary language permits the paraphrase: there are many ways things could have been besides the way they actually are. On the face of it, this sentence is an existential quantification. It says that there exist many entities of a certain description, to wit, ‘ways things could have been’. I believe permissible paraphrases of what I believe; taking the paraphrase at face value, I therefore believe in the existence of entities which might be called ‘ways things might have been.’ I prefer to call them possible worlds. (Lewis 1973, 84)

Lewis claims that we all agree that there are other ways things could have been and that a possible world is just one of the possible ways things could have been. It is not the way we talk about things in ordinary language that commits us to a doctrine about the existence of possible worlds that is difficult to defend. Instead, as Stalnaker states, Lewis’ point is that “…what appears to be a weighty metaphysical theory is really just some ordinary beliefs by another name.” (Stalnaker 1984, 45)

Lewis claims that while we do not always need to take the existential claims of ordinary language at face value, we should do so unless the two following conditions are met: where “taking them at face value is known to lead to trouble”, and where “taking them in some other way is known not to” (Lewis 1973, 84). Lewis defends the claim that positing the existence of possible worlds does not lead to trouble, and maintains that denying the existence of possible worlds does. Lewis

---

52 While Lewis’ defense of his use of possible worlds is central to the issues being discussed in this chapter, his actual theory of counterfactuals is not. Thus, I will focus on the former and forgo any significant explanation of the latter.
argues that rejecting the existence of possible worlds leads to difficulties in explaining modal statements. Linking this to his claims concerning our interpretations of the existential quantifications of ordinary language, he asks, "If our modal idioms are not quantifiers over possible worlds, then what are they?" (Lewis 1973, 85).

Before examining Lewis' position, I note that it is not clear that Lewis' arguments concerning the need to posit the existence of possible worlds are relevant to the use I make of possible worlds semantics. In Chapter Three I claimed that my aim was to find a plausible way to represent to content of the beliefs of non-language-users. Lewis might be right to claim that one who uses possible worlds in order to explain logical necessity and possibility must posit their independent existence. But, if the possible worlds account of propositional content is used only to describe the content of the beliefs of animals, unless one claims that animals have beliefs about logical modality, Lewis' argument may be, strictly speaking, irrelevant.

Stalnaker, of course, applies possible worlds to a wider scope of issues than I do, and argues against Lewis' position on different grounds. I will present Lewis' position on the ontological status of possible worlds in order to clarify Stalnaker's position on the matter. Stalnaker's arguments in favor of the claim that he does not need to adopt Lewis' strong realism about possible worlds provide support for the idea that possible worlds can be used to successfully explain propositional content in all areas of inquiry. I will provide a short explanation of Lewis' positive account of possible worlds, and will comment briefly on his analysis of modal terms. I am most
concerned with his defense of his ontological commitments, since his arguments on these points are directly relevant to Stalnaker's theory, to which I will return shortly.

According to Lewis, there are three alternative options for attempting to explain modal statements of ordinary language that do not depend upon the existence of possible worlds: 1) we could take them to be unanalyzed primitives, 2) we could take them to be analyzable in terms of consistency, or 3) we could take them to quantify not over actual possible worlds, but over sets of sentences which describe different states of affairs or possible ways things could be. I will discuss each of these in turn.

Regarding the first suggestion, Lewis simply says that taking modal idioms to be unanalyzed primitives avoids doing any theorizing at all. We might choose this option if there were no available theories and if modal structures did not seem to be analyzable, but this is not the case. We have, for instance, the theories being considered in this chapter, which analyze modal statements in terms of possible worlds. If a theory that regards modal statements as quantifying over possible worlds is workable, then, unless we find some grounds for rejecting it, it should be adopted over a position that does not produce any workable theory at all. Of course, one reason to reject a theory that depends on the existence of possible worlds might be that its ontology is not acceptable. But this would be rejecting the theory because it leads to trouble. We need to ask whether or not the ontologies of theories such as Lewis' and Stalnaker's lead to problems, but this question differs from the question of whether or not we should attempt to provide an analysis of modal statements, and
needs to be approached independently. While we have promising alternatives to explore, leaving modal terms unanalyzed is not our best option.

The second option Lewis considers is that we take modal idioms to be “metalinguistic predicates analyzable in terms of consistency” (Lewis 1973, 85). In doing so, we take ‘Possibly ϕ’ to mean ‘ϕ is a consistent sentence’ (Lewis 1973, 85). Lewis points out, however, that this leaves “consistency” unexplained. This is a serious problem, since one does not have the option of saying that a consistent sentence is one that ‘could be true’ or one that ‘is not necessarily false’. Doing so would lead to circularity since ‘could be true’ and ‘not necessarily false’ can only mean possible. This causes difficulties for all theories that attempt to define ‘possibility’ in terms of consistency.

An adequate answer to this question is that it does not matter exactly where the problem lies. Lewis’ original claim was only that he had to show that the proposed analysis of modal idioms leads to difficulties. He does show that analyzing possibility in terms of consistency certainly leads to problems. If another analysis of possibility avoids these problems, and does not produce more serious difficulties, we should opt for it over the analysis in terms of consistency.

Another strategy for explaining consistency considered by Lewis understands consistent sentences to be those “…that come out true under some assignment of extensions to the non-logical vocabulary” (Lewis 1973, 85). If we do so, Lewis argues, rather than making a claim that is circular, we are making a claim that is incorrect since, “…some assignments of extensions are impossible, for instance one
that assigns overlapping extensions to the English terms ‘pig’ and ‘sheep’” (Lewis 1973, 85). There will be sentences that will be true only if such impossible interpretations are given. The definition of consistency considered here provides no mechanism for ruling out such cases — we might rule out impossible assignments of extensions on a case by case basis, but then the definition of consistency would not be doing the work. If, on the other hand, we define a consistent sentence as one for which some possible assignment of extensions is true, the problem of circularity arises again.

The third option for analyzing modal terms that Lewis considers is taking them not to be quantifying over possible worlds, but instead to be quantifying over:

...so-called ‘possible worlds’ that are really some sort of respectable linguistic entities: say, maximal consistent sets of sentences of some language (Or maximal consistent sets of atomic sentences, that is state-descriptions; or maximal consistent sets of atomic sentences in the language as enriched by the addition of names for all things that are, that is diagrammed models) (Lewis 1973, 85).

This would allow us to think in terms of possible worlds while at the same time escaping problems associated with references to actual possible worlds, since we are, in fact, referring to entities which are less problematic — sets of sentences. But, Lewis argues, this type of move will result in the same problems that arose from option two above: “...again, the theory would be either circular or incorrect, according as we explain consistency in modal terms or in deductive (or purely model-theoretic) terms” (Lewis 1973, 85). This is because these types of analyses depend on the notion of consistency and lead back to the same difficulties discussed above. If Lewis can show that his explanation of modal statements will work and he can defend
his ontological commitment to possible worlds, then he can provide a non-circular
definition of 'consistency' and avoid difficulties arising from not having such a
definition.

An account of statements of possibility and necessity can avoid the circularity
of the accounts discussed above by reference to possible worlds. In Lewis' formal
system of modal logic, possibility and necessity are logical operators that quantify
over possible worlds. The necessity operator is a universal quantifier over possible
worlds — if a statement is logically necessary, it is true in all possible worlds. The
possibility operator is an existential quantifier over possible worlds — if a statement
is possibly true, it is true in some possible world.\footnote{In fact, there are different types of necessity and possibility. If a statement is physically necessary, it is true in all possible worlds in which the laws of nature are identical to those of our world. If a statement is physically possible, it is true in some possible world that shares the laws of nature with the actual world. Further, we can say of a statement that it is biologically necessary, meaning that it is true in all possible worlds in which the laws of biology are identical with those of our world. Restricting the worlds quantified over by different types of necessity and possibility is dealt with by describing the accessibility relations between worlds. I will discuss only logical necessity, since my point is only to show how it is that reference to possible worlds avoids circularity.} If we explain possibility and
necessity in these terms there is no problem of circularity as long as possible worlds
are entities rather than sets of things themselves defined by 'possibility.' We then
need to explain what a possible world is. Avoiding circularity in the way Lewis
recommends comes at a high cost unless the nature of possible worlds is less
problematic than it has often seemed to be. With this in mind, I turn to Lewis’ view of
the nature of possible worlds.

Lewis says of his view of possible worlds: “When I profess realism about
possible worlds, I mean to be taken literally. If asked what sort of thing they are, I
cannot give the kind of reply my questioner expects: that is, a proposal to reduce possible worlds to something else” (Lewis 1973, 85). He continues:

I can only ask him to admit that he knows what the real world is, and then explain that other worlds are more things of that sort, differing not in kind but only what goes on in them. Our actual world is only one world among others. We call it alone actual not because it differs in kind from all the rest but because it is the world we inhabit. The inhabitants of other worlds may truly call their own worlds actual, if they mean by 'actual' what we do; for the meaning we give to 'actual' is such that it refers at any world \( i \) to that world \( i \) itself (Lewis 1973, 85-6).

'Actual' as Lewis uses the term, is indexical, like 'I' or 'that'; its referent is fixed by the circumstances under which it is uttered. The referent of 'I' changes, depending on which person utters the term. Similarly, according to Lewis, the referent of the term 'actual' changes depending on the possible world in which the term is uttered. From the perspective of any possible world, that world is the actual world, not our world.

Lewis likens his view of 'actuality' to what he claims is a less controversial view of time (Lewis 1973, 86). 'Present' refers to the time at which the term is uttered. When we refer to the present time, we are referring to one time, out of many possible times. Lewis argues that, “We call it alone present not because it differs in kind from the rest, but because it is the time we inhabit. The inhabitants of other times may truly call their own times 'present,' if they mean by 'present' what we do” (Lewis 1973, 86). If we think in terms of an Augustinian view of time, in which only the present is said to actually exist — the future and the past being expectations and memories — 'present' functions very much like 'actual' functions. The future and the
past have an ontological status similar to possible worlds. Each could possibly be the present, and when 'present' is uttered at any time, it refers to that time.\textsuperscript{54}

Lewis' arguments take us from the way we speak about ways things could have been, to his realism about possible worlds. Stalnaker argues, however, that there is a problem with Lewis' argument. Lewis' argument depends on 'possible world' and 'way things might have been' meaning the same thing; if they mean different things, then one cannot be substituted for the other. Stalnaker argues that Lewis' conception of possible worlds commits him to what Stalnaker refers to as extreme realism and claims that extreme realism should be rejected. Nevertheless, he attempts to show that by adopting certain aspects of Lewis' theory, and rejecting others, one can adopt a plausible moderate realism about possible worlds that is both useful and defensible.

Stalnaker discusses four theses defended by Lewis, which Stalnaker claims, taken together, constitute extreme realism. Thesis (1) is that possible worlds exist. Stalnaker explains that thesis (1) claims that "Other possible worlds are just as real as the actual world. They may not actually exist, since to actually exist is to exist in the actual world, but they do, nevertheless, exist" (Stalnaker 1984, 45).

It should be clear, from the presentation above, that Lewis endorses this claim. According to Stalnaker, thesis (1) is compatible with Lewis' claim that believing in possible worlds is nothing more than believing that things could have been different from the way they are. There do seem to be 'ways things might have been.' But,

\footnotesize{\textsuperscript{54} Stalnaker points to Lewis' implicit reference to Augustine. I will return to this topic when I discuss Stalnaker's criticism of Lewis.}
Stalnaker points out, the first thesis says nothing about the nature of possible worlds, only that there are such things.

Thesis (2) is that ‘Other possible worlds are things of the same sort as the actual world — ‘I and my surroundings.’’ Thesis (2) makes a strong claim about the nature of possible worlds. According to Stalnaker, it is the claim “…which gives realism about possible worlds its metaphysical bite, since it implies that possible worlds are not shadowy ways things could be, but concrete particulars, or at least entities which are made up of concrete particulars and events” (Stalnaker 1984, 45).

Thesis (3) is the claim that “The indexical analysis of the adjective ‘actual’ is the correct analysis” (Stalnaker 1984, 45). According to thesis (3), actuality is world-relative; that is, “It is an attribute which our world has relative to itself, but which all other worlds have relative to themselves too” (Stalnaker 1984, 46).55

Thesis (4) is that “Possible worlds cannot be reduced to something more basic” (Stalnaker 1984, 45). This claim is related to Lewis’ arguments concerning whether or not possible worlds might actually be consistent sets of sentences, or some other respectable type of entity. It is important to note that if we could reduce possible worlds to something else, in doing so, we may lose the gains made in avoiding circularity. If possible worlds are entities, we can use them as the foundation on which to stand when we try to explain possibility. Alternatively, if we create them, then we need to build that foundation on something else, and that is difficult to do.

55 For a clear example of Lewis’ support of thesis (3), see the quotation from Lewis above, 110-111.
Taken together, these four theses constitute extreme realism. I claimed earlier that if Stalnaker’s account depends on the actual existence of possible worlds, then his account is false. Lewis, of course, disagrees with this claim. Here, I will present Stalnaker’s arguments in favor of the position he refers to as moderate realism. I will not offer further arguments that Lewis’ theory is false, but assume that, if Stalnaker’s theory can accomplish the same tasks as Lewis’ without taking possible worlds to be independently existing concrete particulars, and defines them in a way that requires ontological commitments that are not as problematic, then Stalnaker’s theory is preferable. I will argue that possible worlds, as explained by Stalnaker, will do the work he needs them to do.

Stalnaker’s position is that theses (1) and (3) should be accepted while theses (2) and (4) should be rejected. The theses he claims constitute Lewis’ account are closely related in many ways. Thus, Stalnaker is faced with two tasks. The first is to show that the four theses are separable — that is, that it is not the case that they must all be accepted or rejected as a group. The second is to show that he is correct about which of the theses should be rejected. I now turn to a discussion of each of the theses, bearing in mind the relationships between the four of them.

Thesis (1) asserts the existence of possible worlds. Thesis (1) says nothing about what type of existence possible worlds have, only that, in some form, possible worlds exist. Lewis’ argument that there are ways things could have been, which are

---

56 There are some parts of Stalnaker’s explanation of thesis (1) that seem to indicate that he takes it to be making a stronger claim than that possible worlds, in some sense, exist. For instance, the claim that “Other possible worlds are just as real as the actual world” (Stalnaker 1984, 45) may seem to imply that there are things that exist but that are not just as real as the actual world. Such statements are far
other than the way things actually are, is adequate support for thesis (1), provided it is not taken to imply anything about the type of existence this entails.

Thesis (2) — that other worlds are of the same sort as the actual world — makes a strong claim about the type of thing possible worlds are. It may seem that thesis (2) simply builds on or adds to thesis (1). But, as Stalnaker points out, if thesis (2) is accepted then 'possible world' must mean something different from 'a way things might have been'. Stalnaker argues that:

If possible worlds are ways things might have been, then the actual world ought to be the way things are rather than I and all my surroundings. The way things are is a property or state of the world, not the world itself. The statement that the world is the way it is is true in a sense, but not when read as an identity statement (Stalnaker 1984, 46).

If 'the way things are' means something different from 'I and my surroundings,' then, 'ways things might have been' means something different than 'possible world.'

While it may follow from the claim that there are ways things might have been — that possible worlds, in this sense, exist — it does not follow that possible worlds are like the actual world. One could accept thesis (1) — possible worlds exist — but still deny that there are possible worlds that have an ontological status similar to the actual world. Similarly, it is not obvious that accepting thesis (1) commits one to the acceptance of thesis (3) — that the indexical analysis of 'actual' is the correct analysis, or of thesis (4) — that possible worlds cannot be reduced to anything else.

outweighed however by Stalnaker's straightforward claims that the first thesis says nothing about the nature of possible worlds (Stalnaker 1984, 45).
While thesis (1) can be accepted even if the other theses are rejected, theses (2) and (3) are difficult to tease apart. It is true, at least, that thesis (2) implies thesis (3). If possible worlds are things of the same sort as the actual world, then Lewis’ claims about the indexical function of ‘actual’ must be correct, since, from the perspective of any world of the same sort as our world, that world is the actual world. If thesis (2) is accepted, there is nothing distinctive about our world, in terms of its actuality, that sets it apart from other possible worlds. If no world is distinct from all others by virtue of its actuality, then ‘actual’ must function indexically.

So, thesis (3) follows from thesis (2). If thesis (2) follows from thesis (3) as well, then the two will of course stand or fail together. It certainly may seem that this is the case, and one might argue that if ‘actual’ functions indexically, then there cannot be anything distinct about any world because of its ontological status. This is because if there were something distinctive, in this sense, about the actual world, then ‘actual’ would apply to that world in a way that is different from the way it applies to all other worlds. If it applies to one world in a way that is different from the way it applies to the others, then it does not function indexically.

Lewis argues along similar lines in his article, “Attitudes De Dicto and De Se.” Here Lewis states, “By ‘possible worlds’ I simply mean certain big concrete particulars, of which this world of ours is one” (Lewis1979, 533). The difference between a theory that takes possible worlds to be concrete particulars and Stalnaker’s, which takes them to be abstract ways things could have been, amounts to the acceptance or rejection of thesis (2). Lewis argues that any theories which take

possible worlds to be abstract entities of some sort will run into problems. He refers to possible worlds on other accounts, which take them to be some sort of abstract entities, as ersatz worlds. Lewis argues that according to any theory that employs ersatz worlds, there will be one such world that corresponds to the actual world. He calls that world the actualized ersatz world. Lewis then argues that on such accounts:

the actualized ersatz world is special, since it alone represents one concrete world. And it is special not just from its own standpoint, but \textit{from the standpoint of any world}. So it is noncontingently special, since contingency is a variation from world to world. But it is part of the theory that the actualized ersatz world is the special one. So it seems to turn out to be a noncontingent matter which of the ersatz worlds is actualized (Lewis 1979, 533).\footnote{While Lewis' arguments here are explicitly directed against Quine, (W.V. Quine, "Propositional Objects," \textit{Ontological Relativity and Other Essays}, (New York: Columbia University Press, 1969) he maintains that although Stalnaker's conception of the nature of ersatz worlds differs from Quine's, since both regard possible worlds as abstract entities, this argument applies equally well to both.}

If it is a noncontingent matter which of the ersatz worlds is actualized, then it does not seem to be possible that 'actual' can function indexically. But beyond this, it needs to be explained how which of the ersatz worlds is actualized could be a noncontingent matter — if possible worlds are in fact possible, then which possible world is actualized should be a contingent matter.

Stalnaker criticizes Lewis' argument and argues that the indexical analysis of 'actual' does not imply that possible worlds are of the same sort as the actual world. First, he accepts the premise of Lewis' reductio and says that, "It is a special fact about "the actual world" (the way things are) that it alone corresponds to the one concrete world" (Stalnaker 1984, 47). However, according to Stalnaker, it does not follow that it is not a contingent fact: "...this is a contingent fact, which means that
from the standpoint of other worlds, it is not a fact. From the standpoint of a
counterfactual possible world, that world has the special property of being the one
and only possible world” (Stalnaker 1984, 47). It might seem as though Stalnaker is
endorsing thesis (2); however, Stalnaker responds that it seems this way only if the
objective standpoint from which to judge all possible worlds is somehow set up as a
neutral standpoint, disconnected from all worlds. He argues that: “...there is no such
standpoint. The objective, absolute point of view is the view from the actual world,
and it is part of the concept of actuality that this should be so” (Stalnaker 1984, 47).

Stalnaker compares the inhabitants of possible worlds to fictional characters.
Just as we take fictional characters to be correct when they, from their points of view,
affirm their own existence, we may also take the inhabitants of possible worlds to be
using language correctly when they say that they are in the actual world. However,
the point of view from a fictional character’s perspective is fictional, and so, “what is
right from it makes no difference as far as reality is concerned” (Stalnaker 1984, 47).
Similarly, what is correct to say in some possible world about what is actual will not
necessarily indicate what is correct to say in the actual world.

According to Stalnaker, it doesn’t follow from the fact that the use of ‘actual’
in other possible worlds to refer to the world in which it is uttered is as correct as its
use in ours, that other worlds are just as real as our own: “the semantic thesis that the
indexical analysis of ‘actual’ is correct can be separated from the metaphysical thesis
that the actuality of the actual world is nothing more than a relation between it and
things existing in it” (Stalnaker 1984, 47). Stalnaker points to the fact that a solipsist
may unproblematically accept the indexical analysis of personal pronouns, and compares this fact to the claim that "...one can accept an indexical analysis of actuality while excluding from one’s ontology any universes that are the way things might have been" (Stalnaker 1984, 47).

Stalnaker expands this explanation with a comparison of his account to the Augustinian view of time. The crucial idea of the Augustinian view, for Stalnaker, is that neither the future nor the past exist, except in our minds as memories or expectations. Stalnaker quotes Augustine’s recommendation that rather than speaking of past, present, and future:

Perhaps it would be more correct to say: there are three times, a present of things past, a present of things present, a present of things future. For these three exist in the mind, and I find them nowhere else: the present of things past is memory, the present of things present is sight, the present of things future is expectation. If we are allowed to speak thus, I see and admit that there are three times, that three times truly are.

By all means continue to say that there are three times, past, present, and future; for though it is incorrect, custom allows it. By all means say it. I do not mind, I neither argue nor object: provided that you understand what you are saying and do not think the future or past now exist. There are few things we phrase properly; most things we phrase badly; but what we are trying to say is understood.

Stalnaker then considers the statement that "...at any time t a future tense statement is true at t just in case the corresponding present tense statement is true at some t' later than t" (Stalnaker 1984, 48). He asks whether this statement is consistent with Augustine’s metaphysical statements about time. It is, Stalnaker argues, as long as

59 I am not concerned here with giving a detailed or accurate account of Augustine’s position. Instead, I wish to present it only as required to demonstrate Stalnaker’s argument concerning the relationship between semantic analysis and ontology.

60 Augustine, Confessions, Book 11, chapter 20, 267, as quoted in Stalnaker (47-8).
"...we understand the times in terms of which our past and future tense statements are interpreted as things which exist in the present" (Stalnaker 1984, 48). This is the way Augustine understands the different senses of time.

If we assume that we accept such a view of time, we can ask questions about how 'now' functions in the past and the future. Only the present really exists on this view, but people in the past and future did, or will, refer to their own times, and in these non-present times, such utterances are true. Stalnaker points out, however, that "...the fact remains (according to our hypothesized metaphysics) that their times do not exist now, which is to say that they do not exist at all (except in memory or anticipation). The fact that those times were or will be real does not show that they are real" (Stalnaker 1984, 48). What is important to notice is that Augustine and those who hold the opposing view — that times other than the present actually do exist in the same way the present does — might share the same semantics for tenses. The key is that the semantic claim can be separated from the metaphysical claim. Even if 'actual' or 'now' function indexically and those who utter the terms in other possible worlds or times use the terms correctly, it does not follow that those other times or worlds actually exist.

This is not the only use that Stalnaker makes of Augustine's conception of time. He also uses a disanalogy between time as Augustine's account describes it and possible worlds as his own account describes them to draw out one of the differences between his position and the strong realist position. The Augustinian claim concerns what is real and what is not; in effect, it posits a boundary surrounding what actually
exists. As Stalnaker puts it, it says "that the actual world is a moment of time"
(Stalnaker 1984, 49). Stalnaker maintains that assertions that the actual world alone is
real do not function in this way. He argues:

…the thesis that the actual world alone has content only if "the actual
world" means something other than the totality of everything there is,
and I do not believe that it does. The thesis that there is no room in
reality for other things than the actual world is not, like the
Augustinian thesis, based on a restrictive theory of what there is room
for in reality, but rather in the metaphysically neutral belief the "the
actual world" is just another name for reality (Stalnaker 1984, 49).

To the extreme realist, who would say instead that if we accept the indexical analysis
of 'actual' then its use in possible worlds other than the actual world picks out some
part of reality, Stalnaker directs the question: "How do I draw the boundaries around
that part of reality which is appropriately related to me to be part of my actual
world?" (Stalnaker 1984, 49). He then points out that any part of reality that is
causally or spatially related to us is part of the actual world.

Other possible worlds are not connected to our world in these ways. If there
are existing possible worlds which are not connected to us in any of these ways, one
needs to ask how we could know anything about them. Stalnaker’s position is that, in
fact, we could not: "If the truth or falsity of our modal claims depends on the
existence of things and events which are causally disconnected from us, then even the
simplest claims about what is possible are unverifiable speculations" (Stalnaker 1984,
49). He explains that his claim is not that there are some modal facts which we can
never know. It is instead that unless there is some kind of relationship between our
world and other possible worlds there is no way that we could know anything about
them. If, instead, we take possible worlds to exist in relation to the actual world and our knowledge of it, we can explain how it is that we can know that what we say is true. This requires that they are not entities of the same sort as the actual world.

These remarks have been intended to show that it is reasonable to accept thesis (3), while rejecting thesis (2). I now turn to thesis (4) — that possible worlds cannot be reduced to some other thing. Thesis (4) is closely related to thesis (2). If thesis (2) is accepted and possible worlds are taken to be of the same type as the actual world, then they will not be reducible to anything else for the same reasons that the actual world is not reducible to anything else. This is Lewis' reasoning when he argues against the claim that possible worlds might be reduced to sets of sentences:

...given that the actual world does not differ in kind from the rest, it would lead to the conclusion that the actual world is a set of sentences. Since I cannot believe that I and all my surroundings are a set of sentences (though I have no argument that they are not), I cannot believe that other worlds are sets of sentences either (Lewis 1973, 86).

I have argued that thesis (2) should be rejected. The rejection of thesis (2) takes away one of the strongest arguments in favor of the acceptance of thesis (4).

The principal motivation to accept thesis (4), once thesis (2) has been rejected, is the idea that if possible worlds are to serve as a foundation for an explanation of propositional content, they need to be solid. If possible worlds are reducible to some other kind of thing, then they seem to be less able to provide a firm base on which to build a theory of content.

Given this, what Stalnaker needs to establish is that even if (4) is rejected, possible worlds can still play the foundational role in his theory that he needs them to.
Once this is accomplished, there are no strong motivations to accept thesis (4).

Without a convincing reason to accept (4), it seems clear that due to the counterintuitive ontological claims that follow from it, and the fact that there is a less problematic ontology with which the same work can be accomplished, it should be rejected.

If we have worries about explanations of consistency, such as those expressed by Lewis, one way out of them is by reference to possible worlds. If possible worlds are independently existing particulars, reference to them is unproblematic. If not, however, then the explanatory power gained from reference to possible worlds depends on our being able to give an account of what they are. Stalnaker says of his theory of possible worlds that:

the possible worlds analysis of proposition yields definitions of consistency and the other propositional relations in terms of elementary set-theoretic relations between the sets of possible worlds determined by the propositions. Whether one accepts these definitions as explanations of the propositional relations may depend on whether, in the end, one accepts the concept of a possible world as having intuitive content and an independent role to play in the theory (Stalnaker 1984, 54).

Whether one finds the concept of possible worlds to be intuitively acceptable depends, in turn, on how Stalnaker’s moderate realism characterizes possible worlds. Although Stalnaker does not provide a reduction of possible worlds to some other entities, he does not see them as standing completely on their own, independent of human activities either:

They obviously are not concrete objects or situations, but abstract objects whose existence is inferred or abstracted from the activities of rational agents. It is thus not implausible to suppose that their
existence is in some sense dependent on, and that their natures must be explained in terms of, those activities (Stalnaker 1984, 51).

One surely may ask, then, what it is about Stalnaker's theory that makes him a realist at all. Stalnaker's explicit answer to this question is that his theory is realist in that:

...it holds that statements about what is possible are to be explained in terms of quantification over possible worlds, and that some such statements are true. It is in the sense that it claims that the concept of a possible world is a basic concept in a correct account of the way we represent the world in our propositional acts and attitudes (Stalnaker 1984, 57).

On the other hand, it is not extreme realism since it "...need not take possible worlds to be among the ultimate furniture of the world" (Stalnaker 1984, 57). Stalnaker explicitly maintains that his conception of possible worlds is not a metaphysical conception. Instead, he takes it to be a "formal or functional notion" and compares its status to that of an individual used in the semantics of extensional quantification theory. He points out that "An individual is not a particular kind of thing; it is a particular role that things of any kind may occupy: the role of the subject of predication" (Stalnaker 1984, 57). He continues, arguing that accepting the semantics for quantification theory does not commit one to any particular metaphysical account of individuals. The same is true of possible worlds:

...a possible world is not a particular kind of thing or place. The theory leaves the nature of possible worlds as open as extensional semantics leaves the nature of individuals. A possible world is what truth is relative to, what people distinguish between in their rational activities. To believe in possible worlds is to believe only that those activities have a certain structure, the structure which possible worlds theory brings out (Stalnaker 1984, 57).
Stalnaker's position is that while possible worlds may not have an existence that is independent of rational activity, possible worlds are what is differentiated between in rational activities. (He does not pause here to define 'rational activity'). Possible worlds theory then, is a way of explaining the structure of, and relationships between, sets of propositions.

This returns us to the question with which this discussion began: Can Stalnaker provide a theory that has real explanatory power while avoiding the difficulties of extreme realism? He certainly avoids extreme realism by his rejection of Lewis' theses (2) and (4). The more difficult question is whether or not his theory is realist enough to provide a tenable explanation of propositional content.

This issue might be raised as follows: In response to the claim that 'A set of propositions is consistent if and only if there is some possible world in which all members of the set are true,' it may be asked, 'Do we need to be appealing to entities that have an existence which is independent from our own if we are to be saying something useful and/or true?' According to Stalnaker, since possible worlds (or ways things might have been) are what we distinguish between in our rational activities, no matter what answers we give to questions about the metaphysical status of possible worlds, not only can such statements be useful, but we can also know that some such statements are true, and others false, even if we do not know exactly what possible worlds are.

Stalnaker argues that his conception of possible worlds is not fundamentally metaphysical. He claims, instead, that what he is doing is describing the structure of
rational activities. This is why he can say things such as, "To believe in possible worlds is to believe only that [rational] activities have a certain structure, the structure which possible worlds helps bring out" (Stalnaker 1984, 57). Such a conception of possible worlds does not rule out strong realism, but it does not require it either. Whether Stalnaker’s non-position concerning the ontological status of possible worlds is successful depends on the acceptability of Stalnaker’s claim that some sentences that quantify over possible worlds are true. Without raising metaphysical questions he can make the claim that:

Anyone who believes that there are objects of propositional attitudes and who accepts the assumptions about the formal properties of the set of these objects, must accept that there are things which have all the properties that the possible worlds theory attributes to possible worlds, and that propositions can be reduced to those things (Stalnaker 1984, 57).

Such a statement is consistent with both realist and non-realist conceptions of propositions and propositional attitudes. To the extent that one is, or is not, a realist about the objects of propositional attitudes, one may also be a realist about possible worlds. The central claim that Stalnaker is making is that his version of possible worlds theory does not require more of an ontological commitment than any other theory of attitudes. If the ontological commitments of the theory are acceptable then acceptance or rejection of the theory should depend on the success of the theory as an explanation of propositional attitudes.

So far, I have argued that Stalnaker’s possible worlds theory provides a plausible account of propositional content and propositional attitudes. It would seem, then, that possible worlds theory provides us with a solid, non-linguistic account of
the content of propositional attitudes. However, there is a problem with possible worlds theory that I have not yet discussed. According to Stalnaker, the objects of belief are sets of possible worlds and the content of a belief is determined by the members of the sets of possible worlds in which the belief is true and the members of the set of possible worlds in which the belief is false. Since logically equivalent propositions are true in the same possible worlds, logically equivalent statements have the same content according to Stalnaker’s possible worlds theory. It follows that for any belief a person may have, he or she believes all logically equivalent beliefs, since, according to the theory, the content of all logically equivalent beliefs is identical. This results in a dramatically counterintuitive consequence pertaining to logically true statements. If a person believes any logically true statement, he or she believes all statements logically equivalent to it, and thus believes all logically true statements. So, any person who believes that 2+2=4, also believes all of the most complex truths of mathematics, since 2+2=4 and all other mathematical truths are true in all possible worlds. However, it just does not seem to be true that anyone who believes in the truth of simple formulas involving the addition of small numbers actually believes all the truths of mathematics. I will refer to this difficulty, as Stalnaker does, as the problem of logical omniscience.

The source of this difficulty seems to rest, in part at least, in the fact that the theory only allows one to deal with possible worlds as complete wholes. Possible worlds, according to Stalnaker’s theory, are not divisible into parts, and content is defined by sets of whole worlds. Capturing the difference between statements such as
‘2+2=4’ and ‘there are an infinite number of primes’ requires some sort of individuation of content that is not possible if content is defined in terms of sets of complete worlds. At the same time, if possible worlds are to do any work at all in the theory, on the surface of it, at least, they must do so as complete worlds. Content is defined, according to possible worlds theory, in terms of sets of complete worlds. If, instead of sets of complete worlds, a theory dealt with parts of worlds, it is the parts that would be understood as defining content, and it is difficult to see how sets of complete worlds could add anything to such an explanation of content. Stalnaker argues, I think correctly, that possible worlds theory captures something crucial to rational activities: the fact that they involve locating the way the world is, within a space of possible ways the world could be, and deciding on actions based on the possible states of affairs those actions will likely bring about. As well, I have argued that Stalnaker effectively argues for a plausible account of the existence of possible worlds. But, his understanding of possible worlds as complete wholes raises the problem stated above and, because of this problem, it is not clear that possible worlds theory gives an adequate account of the content of mental states. In the rest of this chapter, I will examine Stalnaker’s attempts to overcome this difficulty.

Stalnaker has, in several places, attempted to give a satisfactory explanation for the counter-intuitive implication of his theory that logically equivalent propositions are identical, and the further implication that all believers are logically omniscient. Stalnaker’s arguments on this topic have generally been of three types. The first is that there are strong philosophical reasons for explaining belief in terms of
possible worlds, so strong, in fact, that they may warrant accepting some counter-intuitive conclusions. The second is that all theories of meaning suffer from similar difficulties and thus, the difficulty should not lead us to reject Stalnaker’s theory while accepting others. The third is that we can explain the apparent difficulty if we understand beliefs to be about the meaning of symbols rather than about actual propositions. I will address each of these in turn.

In the first of the defenses I referred to above, Stalnaker argues that there is a strong philosophical reason for taking possible worlds to be primitive and for explaining meaning in terms of them. It is central to Stalnaker’s account of belief and desire that they involve distinguishing between alternative possibilities:

It is essential to rational activities such as deliberation and investigation that the participants represent alternative possibilities, and it is essential to the roles of belief and desire in the explanation of action that the contents of those attitudes distinguish between alternative possibilities (Stalnaker 1984, 23).

There are many different forms that the representational system supporting a subject’s ability to distinguish between alternative possibilities could take and different organisms may divide the space of possibilities in different ways. Possible worlds theory is abstract enough to capture all of these, and, Stalnaker argues, strong enough to capture the essential element in representation. That is why, according to him, it makes sense to theorize about representation from the standpoint of possible worlds: “...what any representer must do — what it is to represent — is to locate the world in a space of alternative possible states. It is appropriate to begin with possibilities because that is the level that captures what is essential to representation” (Stalnaker
1986, 115). It is for this reason that Stalnaker claims that thinking about representation should begin with thinking about possible worlds. Even if Stalnaker's theory has some counter-intuitive implications, there are strong philosophical reasons not to abandon the theory, but, instead, to attempt to deal with problems that arise.

As to the second defense, Stalnaker argues that the problem of logical omniscience is a broad-reaching philosophical problem that touches not only his account, but all theories of meaning. He maintains that: “We lack a satisfactory understanding, from any point of view, of what it is to believe that P while disbelieving that Q, where ‘P’ and the ‘Q’ stand for necessarily equivalent expressions” (Stalnaker 1984, 24). Stalnaker points out that the problem concerning logically equivalent propositions is an issue not only for those studying the content of beliefs, but for anyone attempting to provide a theory of deductive reasoning. He argues that this is made clear by any attempt to answer the question of how it is that the conclusions of deductive arguments ever contain any information that is not in the premises. A special case of this is the question of how necessary truths can contain any information at all.

Stalnaker speculates that it may be problematic to see necessary truths as having content at all since we understand statements with content to be differentiating between possibilities:

The difficulty is, I think, that any way of conceiving of necessary truths as having content is at the same time a way of conceiving of them as contingent — as one way things could have been among others. This is, I think, because we do think of content and information in terms of alternative possibilities. Whether the source of my information is my senses, authority, or a faculty of intellectual
intuition with access to a Platonic realm of abstract entities, its deliverances are not news unless they might have been different (Stalnaker 1984, 25).

We take statements about the way the world is to indicate how it is as opposed to other ways it could have been. It can be said in support of Stalnaker's theory, then, that if he is correct in his claims about the source of this problem, even if he does not have a solution at hand, it is a philosophical issue that any complete theory of meaning should address and possible worlds theory gives one a reasonable perspective from which to consider the issue.

If we assume that Stalnaker is correct in arguing that possible worlds semantics is the appropriate perspective from which to approach the problem, we may still ask how the possible worlds theory should treat logically equivalent beliefs that have different content. This leads to Stalnaker's third argument. If two beliefs differ in form, but pick out the same set of possible worlds, then those beliefs have the same content on Stalnaker's theory. So, it would seem that, if a person believes one of those beliefs, he or she also believes the other. What needs to be explained, then, is why someone might assent to the truth of one belief and deny another in cases in which the two beliefs are logically equivalent. Stalnaker sees the difficulty his theory has in dealing with logically equivalent statements as a consequence of the separation of form and content on his theory. Propositions, according to Stalnaker, "...do not have constituents which correspond to the semantically simple constituents of sentences, and do not have an analogue of grammatical structure" (Stalnaker 1984, 23). The content of propositions, according to this view, is not defined by their
constituent parts — since they are taken to have none — but in terms of the external possibilities between which they distinguish; thus, there is nothing in the theory to distinguish between propositions that assign the same truth values to the same possible worlds. Logically equivalent propositions are exactly those propositions that assign truth values to the same possible worlds.

Stalnaker argues that this can be explained, at least in some cases, by the fact that there is a gap between propositions and the statements that express them in such a way that it is not always clear what proposition is being expressed by a sentence. He claims that this will always be the case for non-linguistic accounts of content (Stalnaker 1984, 72). Stalnaker argues that it is not that people knowingly assert and deny logically equivalent statements at the same time. Instead, in such cases, they misunderstand the meaning of at least one of the statements in question. Stalnaker gives the example of a person who believes that \(2 + 2 = 4\) but denies that there are an infinite number of primes. What such a person is lacking, Stalnaker argues, “...is information about the relationship between the sentence ‘There are an infinite number of primes’ and the one necessarily true proposition” (Stalnaker 1983, 73).

According to Stalnaker, what such a person actually believes — since they could not believe that of two equivalent propositions, one is true and the other false — is that the one of the statements means something other than what it does. Stalnaker’s explanation is that:

...whenever a person fails to know some mathematical truth, there is a nonactual possible world compatible with his knowledge in which the mathematical statement says something different than it says in this

132
Learning that the mathematical statement is true is coming to know that that world is not the actual world (Stalnaker 1984, 76).

So, what it is to learn mathematical truths is to come to understand the meaning of mathematical sentences.61

Stalnaker develops this argument further in “Mental Content and Linguistic Form.”62 In this article, Stalnaker distinguishes between coarse-grained content of propositions, which he refers to as informational content — the content of a proposition defined by its truth conditions — and the fine-grained content of beliefs. He explains that one way to describe the problem of equivalence is that the content of beliefs is more fine-grained than informational content.

Here, he asks what kind of information an agent that asserts $P$ and denies $Q$, when $P$ and $Q$ are logically equivalent, must lack. He answers that some of our beliefs are about the informational content of sentences or of our own beliefs. He considers what would have to be true of cases in which $M$ and $N$ are fine-grained objects of belief (whatever such objects may be) that have the same informational content, when a subject $x$ believes $M$ and disbelieves $N$. In cases such as this, $x$ must be ignorant of the fact that they have the same informational content, but then, “...he must be ignorant, either of what content $M$ has, or of what content $N$ has, and this is

61 Stalnaker expresses some doubt about his suggestion, but claims that he has not answered the problem of deduction, but has only shown how his possible worlds account can be shown to be consistent with our belief attributions. Stalnaker also argues that it does not follow from his account that one believes all of the logical consequences of one’s beliefs. This is because, just because we have beliefs, it does not follow that we have compared and combined them all. In effect, one may believe several propositions which, if combined, would imply a conclusion, but since they have not been combined, the conclusion is not believed. See Stalnaker 1984, 85-7, for his explanation of this point and application of it to the problem of deduction.

purely semantic information” (Stalnaker 1999a, 235). This is the type of case — one in which a subject seems to have inconsistent beliefs — that leads us to claim that there must be objects of belief that are individuated more finely than they could be by their informational content. Even in these cases, Stalnaker argues, there is a difference in the information available to the agent concerning the object of belief.

Stalnaker offers the following solution:

Might it be that when we attribute belief that $P$ and deny belief that $Q$ in cases where semantic theory tells us that $P$ and $Q$ are necessarily equivalent, the beliefs being attributed and denied are beliefs that are at least in part about semantic relations? Might the information that distinguishes between necessary truths (which on a straightforward interpretation all have the same informational content) be semantic information — information about the expressions and structures used to state those truths? (Stalnaker 1999a, 235).

This may account for some cases in which a person does not believe informationally equivalent statements. If I believe that I am unmarried, but do not believe that I am a bachelor, I probably do not know what at least one of the terms ‘bachelor’ or ‘unmarried’ means.

Stalnaker considers mathematical beliefs in support of the general idea that the problem of logical omniscience can be explained if we accept that some beliefs are about the content of other beliefs and the meaning of sentences. Stalnaker argues that mathematical information is not completely separable from the form in which it is represented. To support this claim, he considers an imaginary community of English speakers who learn arithmetic in base eight rather than base ten — so they do not use the words “eight” and “nine,” instead “10” and “11” denote the numbers eight
and nine. If a child in the community has a belief he expresses by saying “twenty-six times one-hundred equals twenty-six hundred” Stalnaker asks if it would be correct for us to say that the child believes the twenty-two times sixty-four equals fourteen hundred eight? Stalnaker’s intuition here is that this does not seem to capture the child’s cognitive state and maintains that: “His belief, like our simple arithmetical beliefs, is not really a belief about numbers themselves, independently of how they are represented” (Stalnaker 1999a, 237).

Stalnaker argues further, that our mathematical education and practices almost always involve linguistic or other representations and concludes that: “It is not implausible…to take representations and representational structures to be the subject matter of mathematics, and to be involved in the subject matter in other prima facie cases of necessarily equivalent but distinct objects of belief” (Stalnaker 1999a, 237). If Stalnaker is correct here, then at least part of the difficulty is explained.

Stalnaker’s most recent work on the problem has not offered solutions as much as it has attempted to clarify issues and come to a better understanding of the nature of the difficulty. Stalnaker explores an approach to the problem in “The Problem of Logical Omniscience I”63 that asks why, if one has beliefs, they may not be accessible. The motivation behind this exploration is that if we could explain how it is that one can have a belief, such as one that is equivalent to another belief one accepts, and not have access to that belief, this understanding would help us understand how one could be logically omniscient, yet not seem to be.

Stalnaker is correct that the issue of logical omniscience is a difficult and important philosophical problem, and considering that he provides a strong theory that captures important aspects of rational activities, the fact that he does not have a solution should not, in itself, be seen as a reason to abandon the possible worlds approach. If Stalnaker's arguments that there are strong philosophical reasons for studying meaning from the perspective of possible worlds theory and that the problem of logical omniscience is a problem for all theories of meaning are accepted, then there are good reasons for using the perspective and structures of Stalnaker's theory to address questions concerning representation. This is especially convincing if we also accept that Stalnaker has, in fact, made progress in coming to understand the issues surrounding in the problem of logical omniscience.

As well, while the problem of logical omniscience is a serious issue for possible worlds semantics, as Stalnaker uses the theory, for my uses it may present no problem at all. Stalnaker locates the difficulty not with propositions, but with statements of propositions. It arises, he claims, when there are two logically equivalent statements, one that a subject assents to, and the other that a subject does not. It is difficult to see how such a situation could arise in the case of an animal. So, if Stalnaker has diagnosed the problem correctly, it may be that, while his account faces this problem when applied to humans, the problem does not arise in the case of non-language users, and the theory might be straightforwardly applicable.

Still, if there were a theory that captured what is central to rational activities — distinguishing between alternative states of the world — and which avoided the
problem of logical omniscience, then, assuming it had the explanatory strength of possible worlds theory, the motivation to solve the problem from within possible worlds theory would be lost. Such a theory would be even more attractive if it did not raise questions concerning the ontological status of the key objects referred to by the theory. I will argue in Chapter Five that situation semantics, as presented by Jon Barwise and John Perry, is just such a theory, and that it provides a better foundation on which to base discussions of the content of the thought of non-language-users than does possible worlds semantics.
Chapter Five
Situation Semantics

In Chapter Three, I presented Stalnaker’s position, and argued that the adoption of it could provide a foundation on which to base the attribution of content to the beliefs of non-language-using animals. I argued in Chapter Four, however, that there is a deficiency in the possible worlds approach; according to Stalnaker’s theory, all logically equivalent beliefs have the same content. This leads to difficulties that, I will argue, are better dealt with by a different account of content. In this chapter, I introduce an alternative non-linguistic theory — Jon Barwise and John Perry’s situation semantics — and show how it avoids the difficulty that besets possible worlds accounts, then argue that it can, like Stalnaker’s theory, serve as the basis for a theory of contentful beliefs of non-language-users.64 I stated in Chapter Three that I would present Stalnaker’s theory in an attempt to make progress toward a non-linguistic account of content. I see Stalnaker’s work as valuable in this respect and will rely on it in many ways, and I see his possible worlds theory as one option for providing a non-linguistic account of content. However, I will argue that situation semantics deals with these problems and provides an alternative way of representing the beliefs of non-language-users that is superior to possible worlds semantics.

While Stalnaker offers several reasons why his theory should not be rejected because of the problem of logical omniscience, the only actual solution he presents is

64 The most extensive development of that theory is found in Jon Barwise and John Perry, *Situations and Attitudes* Cambridge by MIT Press 1983. Page numbers in this dissertation refer to the 1999 reprint by CLSI Publications.
another logically equivalent statement, what he or she is making a mistake about is the meaning of the statements, not the content of the proposition that those statements represent. This position is, at best, not fully worked out. However, as I pointed out in Chapter Four, if Stalnaker is correct in locating the difficulty in the interpretation of statements, the identity of logically equivalent statements may not raise problems if the theory is applied only to animals.

As to what I called Stalnaker’s first argument — that his theory captures what is essential to representation and rationality — it is still true that, if another theory could capture the same aspects of rational activity, but avoid the problem of logical omniscience, we should opt for that theory over Stalnaker’s. If such a theory were available, it would also follow that, what I referred to as Stalnaker’s second argument in defense of possible worlds semantics — that all theories of meaning share the problem of logical omniscience — could no longer provide a reason to continue to work with Stalnaker’s theory. A theory that escaped the problem of logical omniscience, but accomplished all of the same work as Stalnaker’s theory, would clearly be superior.

I am going to argue that a version of situation semantics will, in fact, accomplish these things. I will describe situation semantics as it is presented by Jon Barwise and John Perry, then I will discuss it in relation to Stalnaker’s theory, and finally, I will argue that situation semantics can be combined with Bennett’s work in the same way that I argued Stalnaker’s could in Chapter Three.
According to Barwise and Perry, reality is made up of situations that are related to each other in various ways. A situation is what one might intuitively think of as a state of affairs or course of events constituted by objects and the relations between them. The basic idea behind situation semantics is that meaning can be explained in terms of the relationships between situations. Although situations are related to each other in very complex ways, each situation is unique. Living organisms, then, constantly find themselves in unique situations. In order to survive, many animals must not only deal constantly with new situations, but, in many cases, in order to act appropriately they must also have the ability to anticipate the direction courses of events will take. This requires the ability to gain information about some situations from the observation of other situations. To use examples from Barwise and Perry, a smell, shadow, or sound may alert an animal to the presence of a predator or prey, and smoke pouring from a window may alert one to the fact that one’s house is on fire. This recognition of the similarities between situations is key to successfully maneuvering through one’s environment. Barwise and Perry say that it is true of any organism that depends on acting appropriately in changing environments that:

it is by categorizing situations in terms of some of the uniformities that are present, and being attuned to appropriate relations that obtain between different types of situations, that the organism manages to cope with the new situations that continually arise” (Barwise and Perry 10).

It is out of this need to classify and anticipate situations, argue Barwise and Perry, that meaning arises. In giving a more detailed account of their theory, I begin with a
general overview of the ideas behind situation semantics. I then turn to a short
presentation of some of the more formal aspects of the theory. After that, I describe
how Barwise and Perry represent the content of mental states in their theory. Then I
explain how they avoid the problem of logical omniscience. Finally, I will argue that,
like Stalnaker’s work, Barwise and Perry’s can be consistently combined with
Bennett’s.

Much of the nuts and bolts work of the theory is in the classification of
situation-types and analysis of how such types are related to each other. Since my
concern is with the fundamental aspects of the theory and the possibility of it
providing an example of how the content of the beliefs of non-language–users can be
understood, I will not discuss the technical aspects of the theory in detail, but instead,
I will focus on its foundations in order to show how it provides an approach different
from Stalnaker’s and how this approach avoids its difficulties. Still, I need to
introduce some of Barwise and Perry’s technical terminology and notation in order to
explain the foundational aspects of their theory.

Situations, on this account, are described in terms of their components —
individuals, properties, relations, and locations. These components are, in Barwise
and Perry’s terminology, uniformities across situations. That is, any component of
this sort could be a part of various situations.

Being able to infer information about one type of situation from another
depends on those situations being related in a systematic way:

One situation $s$ can contain information about another situation $s'$ only
if there is a systematic relation $M$ that holds between situations sharing
some configuration of uniformities with $s$ and situations that share
some other configuration of uniformities with $s'$ (Barwise and Perry
14).

The uniformities may be physical or abstract objects of the type mentioned above.

The key idea is, of course, that situations must be systematically related. Meaning
arises through the recognition and exploitation by organisms of the relations between
situation-types. As Barwise and Perry put it, "...the various uniformities across
situations (objects, properties, etc.) and the various relations between them that give
rise to meaning stem directly from an organism's need for efficient ways to cope with
the stream of situations it encounters" (Barwise and Perry 14). For Barwise and Perry,
since meaning is a product of the relationships between situations, to study meaning
is to study such relations as they are significant to us, or to other organisms that
utilize the systematic relationships between situations in order to cope with their
environment.

The meaning of mental states is explained by the theory in terms of how
mental states, as situations, are systematically related to other situations in the world.
Barwise and Perry provide an example: Molly the dog wags her tail when she hears
the voice of any member of the immediate family of her owner. When she hears the
voice of any person who is not a member of the family of her owner, she barks as
though she is being attacked. For Molly then, there is a property of people that might
be called 'pro' and another that might be called 'con'. These properties are
systematically linked to Molly's perception of people and her actions when she hears
their voices. Molly is reacting differently to different situation-types. But, Barwise
and Perry argue, considering the situations in the world is only one side of what we
should be concerned with:

The other side takes place in the dog. Molly goes into one state, say *pro*, when she perceives pros, and another, say *con*, when she perceives cons. These states are uniformities across situations involving the dog, just as properties of *pro* and *con* are uniformities across the situations the dog confronts. The properties and corresponding states are like two sides of the same coin the dog confronts (Barwise and Perry 12).

Barwise and Perry’s answer to the question of why it is that perceptual states mean what they do is that there is a systematic relation between such states and states of the world. Their approach involves classifying the relations between subjects and the world rather than attempting to discover what it is about subject’s internal states that make it true of them that their inner states are intentional. Molly’s *pro* and *con* states are described not by reference to her brain states or to any inner mental objects, but instead, by a description of the classes of objects to which she reacts in systematic ways.

According to Barwise and Perry, the same is true of descriptions of the mental states of humans. They consider Sarah, a human, watching a baseball game on television. After explaining her beliefs about which team will win, who will strike out, etc., Barwise and Perry point out that they “...managed to describe Sarah’s mental state by referring not to anything in Sarah’s circuitry or inner world, but to people, places, properties, and relations miles away from Sarah, things that could not possibly be in Sarah’s head” (Barwise and Perry 225).
Barwise and Perry claim that describing a subject’s mental state in this way depends on indirect classification. What they mean by ‘indirect classification’ is illustrated by an example they give of two aerosol cans, one containing bug spray, the other hair spray. We might place a label on one with ‘BUG’ and a label on the other with ‘HAIR’. There is no problem in choosing which one to use on one’s hair and which one to use to rid one’s kitchen of roaches. The labels do not tell us what is in the cans. Neither can has bugs or hair in it. As Barwise and Perry put it, “We are exploiting the nomic constraint between spraying bugs with a certain chemical and dying bugs to classify the chemical by referring to bugs. Similarly with the hair spray” (Barwise and Perry 227).

At bottom, when it comes to describing the mental, Barwise and Perry attempt to classify situations and document the ways in which situations and attitudes are systematically related. Barwise and Perry take mental attitudes to be bodily states standing in relation to situations in the world, and, as such, to be themselves situations: “Beliefs, thoughts, and perceptions are situations (including both states of affairs and events that include activities) involving an agent at a location being in bodily conditions” (Barwise and Perry 230). According to their theory, these states are indirectly classified by attitude statements.

Barwise and Perry’s indirect classification of attitudes depends on being able to describe attitudes in terms of the situations in the world to which they are regularly related. The account is far more straightforward in cases concerning perceptual attitudes — seeing, hearing, or otherwise perceiving x — than it is for intentional
states that are less directly linked to situations in the world. Still, the classification of beliefs and other non-perceptual attitudes is developed in roughly the same way. Barwise and Perry take beliefs to be dispositional states in the sense that they are “…real states known through their effects” (Barwise and Perry 241). By dispositional, then, they do not mean to say that beliefs are reducible to behavioral patterns. Instead, their claim is that, “…belief states are efficient bodily conditions (about which virtually nothing is known) that have enough structure to account for the complex information-retaining and behavior-guiding capacities that we use them to explain” (Barwise and Perry 242). Beliefs then, on this theory, are indirectly classified in terms of their effects.

According to this account, mental attitudes do not have propositions as their objects. In fact, Barwise and Perry explain that in Situations and Attitudes, they deliberately avoid the terms ‘proposition’ and ‘propositional attitude’ and hold that propositions are the artifacts of semantic theories and “…not the sorts of things people usually talk about, even when reporting someone’s attitudes” (Barwise and Perry 178). Barwise and Perry do not claim to be describing the objects of attitudes at all. What they are doing instead is describing the attitudes of subjects indirectly, by describing those subject’s relationships to situations. Ultimately, for them, this amounts to describing relationships between situation-types, some involving subjects in the way that attitude reports apply to them, and some involving situations in the world.
Before turning to a comparison of Barwise and Perry's work to Stalnaker's, I will describe some of the more formal aspects of situation semantics. This will provide a more accurate picture of how mental attitudes are characterized within the theory. My description of these is brief and does not do justice to the complexity and subtlety of Barwise and Perry's theory. I have only included enough of a description to indicate how their approach differs from Stalnaker's how it can be used in combined with Bennett's.

Situations are constructed, according to the theory, out of individuals, relations, properties and locations. Situations are actual states of affairs or courses of events in the world. Although each situation is unique, different situations can be classified as being of the same type. For example, a situation in which Molly barks and a situation in which Jackie barks are both situations in which a dog barks. In this respect then, they are of the same type.

A situation-type is, “a partial function from n-ary relations and n individuals to the values no (‘false’) and yes (‘true’)” (Barwise and Perry 8). That is to say that a situation-type indicates a value of ‘yes’ or ‘no’ for each of the claims which is it defined in it. In the notation of the theory, the situation-type s, in which the dog, Molly, barks, the dog, Jackie, does not, and the dog Spot is not present, is described by:\(^\text{65}\)

\[
\begin{align*}
\text{In } s: & \text{ barks, Molly; yes} \\
& \text{ barks, Jackie; no} \\
& \text{ barks, Spot; undefined}
\end{align*}
\]

\(^{65}\text{See Barwise and Perry 8-9 and 54 for their explanation of this notation.}\)
If a situation is of type s, then it is the case that, in that situation, Molly barks and that Jackie doesn't bark. Any situation in which Molly does not bark, or Jackie does bark, will not be of type s. As well, one situation can be of many different types at the same time. For example, at the same time a situation is of the type in which a dog is barking, it may also be of the type in which a beagle is barking, or of the type in which a beagle is left alone, or is hungry.

What Barwise and Perry refer to as ‘courses of events’ are also central to the theory. A course of events is “a function from locations to situation-types” (Barwise and Perry 1983, 9). Courses of events describe occurrences at one or more locations. For Barwise and Perry, a location is a general region of space and time — a place in which an event can occur, rather than a precise geometric point. A course of events, defined for a particular location, indicates a situation-type. Barwise and Perry use e, e', e"...to range over courses of events. A course of events e, that includes Molly barking at location l, Mr. Levine shouting at Molly at location l', and Molly not barking at location l"", can be represented as follows:

In e, at l: barks, Molly; yes
at l': shouts at, Mr. Levine, Molly; yes
at l"": barks, Molly, no

A course of events indicates a situation type for locations in which it is defined. In this example, Molly barks at one location, Mr. Levine shouts from another, and Molly stops barking at a third. The notation does not indicate how close Mr. Levine is to Molly in physical distance or time. 66 All of the coe’s I refer to, with the exception of

---

66 See Barwise and Perry, 9 for their explanation of these examples.
the one immediately above, consist of only one event. While individuals, properties, relations, and locations are taken by Barwise and Perry to actually exist, to be part of the furniture of reality, to borrow Stalnaker’s language, situation types and courses of events are not. Instead, they are abstract, set-theoretic properties.

Barwise and Perry also make use of what they call ‘event-types’ in order to capture, within their theory, similarities between different courses of events. $E$, $E'$, $E''$... are used to represent event types. Event types are treated just like courses of events and, in fact, are just like courses of events, except for the fact that within event-types, “…abstract proxies may appear in place of genuine individuals, genuine relations and genuine locations” (Barwise and Perry 71).67 Barwise and Perry refer to the abstract proxies they use as ‘basic indeterminates’: while situations are made up of individuals, relations, and locations, event types may include basic indeterminates as well as these things. There are three types of basic indeterminates to which I will refer: individual indeterminates standing for individuals, relation indeterminates standing for relations, and location indeterminates standing for locations. I will use $a$, $b$, ... to represent individual indeterminates, $r$, $s$, ... to stand for relation indeterminates, and $l$, $l'$, ... to refer to location indeterminates.68 An example of an event type is one

---

67 Barwise and Perry explicitly distinguish real situations from the abstract situations they use to classify real situations. Abstract states of affairs, courses of events and event-types are sets. Because they are abstract, they have no causal properties. Real situations are, on the other hand, parts of the natural world; they are perceived by the senses, and are caused by and cause other events (Barwise and Perry 57-8). Barwise and Perry argue further that because they use abstract objects only to organize and classify real situations, they avoid difficulties relating to the ontological status of such objects (Barwise and Perry 72).

68 These symbols differ slightly from those used by Barwise and Perry in Situations and Attitudes.
that consists of someone being a tired and hungry philosopher. We can represent this event-type as follows:

\[
\text{in } E: \text{ at } I: \text{ philosopher, } a; \text{ yes}
\text{ tired, } a; \text{ yes}
\text{ hungry, } a; \text{ yes}
\]

Obviously, there are many situations of this type. Each may involve different philosophers who are tired and hungry and may be at different locations (Barwise and Perry 73). Basic indeterminates, like event-types and situations-types, are set theoretic entities that are purely abstract (Barwise and Perry 71).

Barwise and Perry call the linking of event-types to real situations 'anchoring'. A function \( f \) assigns real individuals, relations, and locations to the indeterminates in an event-type. Barwise and Perry refer to such functions as 'anchors'. If each indeterminate in an event-type is assigned a value by an anchor, then a particular course of events, which is of that type, is indicated. In such a case, the event-type is said to be anchored to the event.

Barwise and Perry employ these tools in their explanation of mental attitudes. Much of the work they do is motivated by issues, and designed to overcome problems, concerning the semantics of attitude reports. They show how the meaning of sentences that report perceptual and belief states can be explained and represented. This project, in the end, involves giving an account of mental attitudes themselves.69 All of this work falls under a general account of meaning; still, it is important to keep in mind the difference between descriptions of attitude reports and descriptions of attitudes themselves.

69 See Barwise and Perry 220-3 for their comments.
Barwise and Perry begin their discussion of attitudes with an explanation of perceptual states, and then extend this account to belief states. In discussing reports of perceptions, Barwise and Perry draw a distinction between attitude reports that are epistemically neutral and reports that are epistemically positive. The statement 'Bob saw Jack stealing a car' might be true in some circumstances, even if Bob did not come to believe that Jack had stolen the car. For example, Bob may have seen Jack steal a car while thinking that Jack was getting into Jack's own car and driving away. In this case, the sentence is epistemically neutral. On the other hand, the statement, 'Bob saw that Jack was stealing a car' is only true if in fact Bob realizes that what he was witnessing was a theft of a car, and that it was Jack who is doing the thieving. What must be the case in order for the epistemically neutral sentence 'Bob sees x' to be true, is that Bob's eyes must be open, Bob must be alert and have the perceptual capacity to see, and x must be in Bob's field of vision. This tells us very little about the information Bob takes in; instead, it tells us primarily about what is in Bob's field of vision. In terms of situation semantics, the emphasis of epistemically neutral perceptual reports is with the scene observed rather than the state of the observer.

In the case of epistemically positive perceptual reports, on the other hand, there is information given both about the scene observed and about the state of the observer. At any location, there are a number of scenes, some observable and some not. As well, from different positions of observation, different scenes are observable. Statements that are epistemically positive report on what information the observer gains from perceiving a situation about the way the world is. Barwise and Perry
explain that epistemically positive reports are appropriate when an agent stands in relation to a collection of alternative situations that are compatible with what she sees and knows (Barwise and Perry 177). In such cases, it is often important to indicate which of the possible pieces of information that could have been gained, were gained. Reports that a being sees that something is the case, in this sense, classify a state of the perceiver by indirectly referring to his or her inner states that indicate which features of the observed scene correspond to those inner states. The idea is that organisms perceive uniformities in situations and that there are features of organisms' inner states that correspond to the perceived features of the environment. What Barwise and Perry do is use event-types to refer to both the features of the environment that are perceived and the perceptual state of the observing organism.

The point of the distinction between seeing reports and seeing-that reports is not that some animals have the capacity to see scenes, but not to see that something is the case. The distinction is between different types of statements used in different types of circumstances. ‘Seeing that’ statements are used when there are alternative possible situations that might have been seen by one who observes a scene. For example, a student comes into my office and sees a pile of papers on my desk. It is a further question whether a particular student who walks into my office sees that one of the papers on my desk is the test that will be given tomorrow rather than seeing only that my desk is messy. Barwise and Perry are not trying to distinguish between two different types of seeing, but between two different types of reports. My interest

70 See my discussion of Norman Malcolm’s position, in Chapter Six, for an example of an explanation of attitude reports that does draw this type of distinction.
is in questions concerning the attribution of content to mental states. Because epistemically positive reports refer to perceptual states by virtue of the content of those states, I will focus on Barwise and Perry’s discussion of them, rather than on epistemically neutral reports.

Describing perceptual states involves describing which aspects of a perceived scene an organism takes in as information. Epistemically positive perceptual reports do just this. Accordingly, Barwise and Perry’s account of perceptual states grows out of their account of epistemically positive perceptual reports. Barwise and Perry use seeing as an example of how perceptual states can be described.

Their explanation begins with an explanation of the relation $ST$ (sees that) holding between observers and events. Using the components of the theory described above, an example of the $ST$ relationship is as follows:

in $e_o$: at $l$:ST, $a$, $e$; yes.

According to this statement, the event $e$ is part of the event $e_o$ in which, at location $l$, agent $a$ sees that $e$. But, the event $e$ can be classified as being of various types, and this does not capture what the event $e$ is perceived as.

Barwise and Perry remedy the problem by providing a way of identifying the event-type that an event is observed as. The perception of an event can be represented by linking the actual event to an abstract event-type, thereby isolating the features that are perceived. As Barwise and Perry explain: "...given event-type $E$ and a function $f$ that anchors $E$ in $e_o$ such that $E[f]$ is part of $e$, we can use $E$ and $f$ to represent aspects of the perception" (Barwise and Perry 234). The anchored event-type is part of the
observed event — it is that part of the event that is observed by the agent. (One event is part of another event if whatever occurs in the first, occurs in the second.) Thus the event and event type indicate what is observed, and what it is observed as. As Barwise and Perry explain:

The event-types exploit the duality involved in perception; they classify a perceptual system as being in a condition suited for perception of a certain kind of environment, from a certain location in it, and an environment as suitable for a certain kind of perception. The role of indeterminates is to coordinate these two complementary uniformities (Barwise and Perry 234).

In these cases, the agent’s inner-state, or perception, is characterized as being of a certain type. Barwise and Perry refer to the relation there described as represented perception (R-perception) and use $S_r$ (R-seeing) to refer to the relation that holds between agents and event-types in such cases.

Barwise and Perry provide an example of a description of Joe’s seeing Jackie biting Molly. This event can be described as an R-perception of an event type:

\[
\text{in } e: \text{ at } l: S_r, \text{ Joe, } E; \text{ yes.}
\]
\[
\text{of, } t', \text{ Jackie; yes}
\]
\[
\text{of, } t'', \text{ Molly; yes}
\]

where

\[
E: = \text{ at } k: \text{ biting, } t', \ t'' ; \text{ yes.}
\]
\[
\text{dog, } t'; \text{ yes}
\]
\[
\text{dog, } t''; \text{ yes}
\]

Here the event is described in terms of Joe’s R-perception of the event-type in which one dog bites another, that is anchored by Jackie and Molly as the participating dogs. The first part of this description — ‘in $e$: at $l$: $S_r$, Joe, $E$; yes.’ — indicates Joe’s inner state or perceptual condition, that of perceiving event-type $E$. It indicates Joe’s inner state, by indicating the type of event Joe perceives. Barwise and Perry refer to the
second part — 'of, t', Jackie; yes' and 'of, t'' Molly; yes' — as the setting. It links that event type to the actual event that is perceived. There are many different events of type E, events in which one dog bites another. Describing perception in this way indicates what is common to all such events. But also, by anchoring the event to Jackie and Molly, the description gets at what is particular to the event.

Barwise and Perry claim that attitude reports in folk psychology assume that perception should be described as having these two parts:

...the use we make of attitude reports in Folk Psychology presupposes this bipartite classification of perception. The uniformities across individuals captured by the constraints of Folk Psychology operate at a level of frame of mind. But the frame of mind alone loses track of the way the perceiver sits in the world, what has caused the perception and, hence, what the perception is of (Barwise and Perry 236).

If we only describe Joe's inner state, they claim, we do not capture this perceptual event.

Joe could in fact be in the same state of mind if he saw another dog biting his dog, Molly: To keep track of what he saw, we need to anchor his perceptual condition to the objects in the world that caused it. This is the reason we must factor the perceptual event into two parts that fit together, the frame of mind and the setting in which the perceptual event takes place.

Barwise and Perry build on their work on perception in order to provide a way of describing belief. They take beliefs to be complex event-types that have ideas as constituents (Barwise and Perry 242). Unlike perceptions, ideas are not necessarily of any existing thing. Still, Barwise and Perry describe beliefs as having two parts. In the case of perception, the relevant frame of mind is the perceiver's perceptual
condition. In the case of belief, Barwise and Perry refer to a believer's doxastic condition. As in descriptions of perception, in descriptions of belief, the setting is also important; this is not necessarily the place the believer is in at the time of having a belief, but the objects in the world to which the belief can be said to be about, if there are such objects. They classify agents' doxastic conditions by describing the relation $B_r$ (represented belief) that holds between agents and event-types.\(^1\) Joe's belief that Jackie is biting Molly can be represented as:

\[
\text{in } e_o: \text{ at } t: B_r, \text{ Joe, } E; \text{ yes.}
\]
\[
\text{of, } a', \text{ Jackie; yes}
\]
\[
\text{of, } b'', \text{ Molly; yes}
\]

where

\[
E: = \text{ at } h: \text{ biting, } a', b''; \text{ yes.}
\]
\[
\text{dog, } a'; \text{ yes}
\]
\[
\text{dog, } b''; \text{ yes}
\]

Here we have Joe's belief that an event of the type in which one dog is biting another and in which the dog doing the biting is Jackie and the dog being bitten is his dog Molly. The first line, which claims that the $B_r$ relation holds between Joe and event-type $E$ refers to Joe's frame of mind — the state that Joe must be in if he believes that $E$ is occurring. Barwise and Perry make no attempt to explain what this condition must be like in physiological terms or how we might determine whether or not Joe is in such a state. They assume, as indicated above,\(^2\) that beliefs are internal, bodily, states

---

\(^1\) Barwise and Perry use sets of indexed event-types (or schemata) rather than event-types because they are needed in order to describe complex beliefs. However, Barwise and Perry use event-types in their examples and referring only to event-types is sufficient for my goal explaining their fundamental ideas as well.

\(^2\) See above, 145.
known through their effects, and leave it an open question as to how such effects might be linked to beliefs.73

Barwise and Perry, then, have a theory according to which meaning is explained as existing due to the relationships between situations. They use the theory to explain the content of mental attitudes in terms of the situations to which they are systematically related. As well, the theory provides an account of how the content of such attitudes can be represented in terms of the events and event-types as explained by the semantics of the theory. I now turn to a discussion of some important differences between Barwise and Perry’s account and Stalnaker’s. In particular, I will describe two important ways in which in which the two theories disagree, and some of the implications of these differences. First, I will consider what each theory takes as primitive, and second, I will consider the fact that while possible worlds are taken to be complete, situations are not.

The types of things the theories of Barwise and Perry and Stalnaker take to be primitives — the fundamental building blocks out of which the structures of possible worlds theory and situation semantics are built — differ significantly. In situation semantics, the primitives of the theory are the parts that make up situations — individuals, properties, relations, and locations. These are all entities that, while perhaps giving rise to metaphysical questions, are generally taken to exist in some meaningful sense. As Stalnaker puts it in a discussion of situations, the primitives of

73 I do not mean this as a criticism of Barwise and Perry’s work at all. They are simply more concerned with how to represent beliefs once it is known what beliefs a subject has. This is certainly a big enough project in itself.
Barwise and Perry's theory are "...things that actually exist, and that exist in the natural order, independently of thought." Stalnaker's theory on the other hand, takes possible worlds to be primitives. Properties and relations are defined in terms of possible worlds and individuals. Taking possible worlds as the starting point of the theory has some serious implications that might incline one to adopt the situation semantics approach to explaining attitudes. I will return to this issue shortly.

A second difference, and serious dispute between Stalnaker and Barwise and Perry is that Stalnaker takes possible worlds to be complete, in the sense that, within a possible world, all statements concerning the objects in the world have a truth-value, while Barwise and Perry's theory allows for situations with partial information content. Thus, on Stalnaker's account, propositions are understood to be total functions; that is, they indicate a value of T or F for every relevant possible world. The same is not true of situations. According to Barwise and Perry's theory, situations are partial in the sense that there will be statements concerning their objects for which the truth value will be undefined.

As discussed earlier, Stalnaker's theory treats logically equivalent statements as having the same content. So, on Stalnaker's account, the mathematical statements ‘2 + 2 = 4’ and ‘There are an infinite number of primes,’ have the same content since they are true in the same sets of possible worlds.

---

75 I say every relevant possible world since, according to Stalnaker, possible worlds are the product of rational activities, and as such, in any inquiry, only those possible worlds relevant to that inquiry need be considered. Thus, what counts as the set of all possible worlds is only the set of all worlds relevant to the inquiry at hand. I consider this issue in more detail in Chapter Six.
Situation semantics treats such statements differently. One key advantage to Barwise and Perry's approach is that, according to it, logically equivalent sentences need not carry the same informational content. Barwise and Perry explain logical equivalence in terms of models. Sentences that are true in all of the same models are logically equivalent. A model is an assignment of appropriate set theoretic entities to the terms and relations of the language involved" (Barwise and Perry 25). Barwise and Perry consider the following two sentences in explaining how their theory differs from others, like Stalnaker's:

1) Joe is eating.

and

2) Joe is eating and Sarah is sleeping or Sarah is not sleeping.

If we assign appropriate set theoretic entities to all of the terms in each of the two sentences, the truth of both of the sentences, in any model, will depend only on the truth of the assertion that 'Joe is eating' since the assertion that 'Sarah is sleeping or Sarah is not sleeping' will be true in all models. Thus, the models in which the sentence (2) is true are precisely the models in which sentence (1) is true and the sentences are logically equivalent.

As explained in Chapter Four, according to Stalnaker's theory, logically equivalent statements will be true in the same set of possible worlds, and since content is also determined by the set of possible worlds in which a statement is true, according to Stalnaker, logically equivalent statements have the same content. In
Barwise and Perry’s theory, however, sentences are taken to stand for situations. Sentence (1) stands for situations in which Joe is eating, including all situations in which Sarah is sleeping or Sarah is not sleeping. Sentence (2) stands for only those situations in which Sarah is present, awake or asleep, and Joe is eating.

Barwise and Perry explain that developing a system that captures the partial nature of information was one of the key motivations behind the development of situation semantics. Their particular interest was in explaining naked-infinitive perceptual reports such as ‘Sarah saw Joe eating.’ In reports of this grammatical structure, the verbs in the embedded sentences — in the case of the example, ‘Joe eating’ — are not conjugated as they normally would be. A standard explanation of this is that the action to which the verbs in the embedded sentence refers is taken to occur at the same time of the act, and the verb referring to the perception indicates, through its conjugation, the time of the action.

When Sarah sees that Joe is eating, she sees something, which Barwise refers to as a scene, which contains some information, but not all information, about Joe, his surroundings, or the weather in Alaska. Barwise and Perry argue that intuitively, what Sarah saw should described by a theory that respects this. In possible worlds semantics the basic idea is that there is a set of possible worlds — each complete and determining the truth value of all propositions concerning the objects in that world.

---

76 Actually, this is slightly misleading. Barwise and Perry are quite concerned with the considerations relating to the context dependence of meaning and the fact that the same sentence can be used to stand for a variety of situations. But, the claim that sentences stand for situations is close enough to the truth to be useful in explaining the difference between situation semantics and possible worlds theory. (See Barwise and Perry 1983, 24 on this point).

The worlds in which the truth value of 'Sarah saw Joe eating' is true are the worlds in which Sarah saw Joe eating. The things that Sarah did not see, and all other aspects of the world not necessarily related to Sarah's seeing Joe eating, are things that are not common to all of the worlds in that set. Barwise and Perry criticize this strategy as being cumbersome and unintuitive and, of course, for the result that those whose beliefs are being described — in my example, Sarah — know all logical truths. (Barwise and Perry 1999, xxiv). Thus, situations differ from worlds in that situations are partial and provide answers to the truth values of some, but not all, statements.

Understanding propositions as complete functions in Stalnaker's theory is a product of the fact that he accepts the consequences of explaining content in terms of possible worlds. If the idea that possible worlds are worlds is taken seriously, then even if they are abstract products of rational activities, possible worlds are taken to be fully detailed. So, in the case of any particular world, there is an answer to any question concerning any object in that world. Since Stalnaker starts with possible worlds and takes propositions to be functions from possible worlds to truth values, propositions are, for him, total functions.

Taking possible worlds to be complete leads to the problem of logical omniscience. This can be seen to be the case since, as Perry argues in, "From Worlds to Situations," if partial functions were admitted to Stalnaker's theory, the problem of logical omniscience could be avoided. Reviewing this argument brings out

important differences between situation semantics and possible worlds theory clearly.

For use in his argument, Perry defines the term 'issue' in the following way: "An n-adic relation $R$ and appropriate objects $a_1, \ldots, a_n$ determine a basic issue, the world is such that $R$ holds of $a_1, \ldots, a_n$ or not. A basic issue has an answer: yes or no" (Perry 85). Possible worlds, or ways the world could be, Perry refers to as ways, or total ways, since they provide answers to every issue under consideration. It is suggested that Stalnaker's theory could be extended to include partial ways in addition to total ways. While total ways provide answers to all issues under consideration, partial ways provide answers to only some such questions. Partial ways, then, are just parts of total ways.\(^{79}\)

In addition to total functions, Perry suggests that Stalnaker's theory could recognize a wider class of propositions that includes partial functions from total ways and partial ways.

For example, on the issue of whether George is sleeping, the type of function Perry is advocating will be undefined in certain possible worlds, or ways. The partial function $P$ from ways to truth-values on the issue of whether George is sleeping is such that:

\[
P(w) = T \text{ if } w \text{ provides an answer yes to the issue of whether George is sleeping.}
\]

\[
P(w) = F \text{ if } w \text{ provides the answer no for that issue.}
\]

\(^{79}\) Perry expresses some concern over distorting the spirit of Stalnaker's theory by describing possible worlds as taking us from issues to answers, since Stalnaker takes possible worlds to be primitive, and seeing them in terms of issues to answers, takes them to have parts. However, he argues further, that he sees no reason Stalnaker should not do so (Perry 85-88).
\( P(w) \) is undefined, if \( w \) provides no answer for that issue.

The advantage that Perry claims comes from this approach is that propositions that are necessarily equivalent will not always turn out to be identical. Take the function \( P' \) to be the proposition that George sleeps and Mary weeps or Mary doesn’t weep.

The function \( P' \), can be described:

\[
P'(w) = T \text{ if } w \text{ provides an answer to yes to the issue of whether George is sleeping, and provides the answer yes or no to the issue of whether Mary is weeping.}
\]

\[
P'(w) = F \text{ if } w \text{ provides the answer no to the issue of whether George is sleeping, and provides the answer yes or no to the issue of whether Mary is weeping.}
\]

\( P'(w) \) is undefined otherwise.

\( P' \) is logically equivalent to \( P \), but the functions are not identical. The set of worlds in which \( P \) is defined is not the set of worlds in which \( P' \) is defined.

Similarly for mathematical sentences, if one accepts that there can be partial ways and partial functions, then statements like ‘2 + 2 = 4’ and ‘There are an infinite number of primes’ have different contents and the theory can attribute to a person one belief while holding that the person does not believe the other.\(^{80}\)

Barwise and Perry argue that Stalnaker’s theory could be expanded to account for partial functions. Stalnaker’s theory would then be changed only by including partial functions and worlds. Perry argues that this would, in fact, make it a version of situation theory, and that since nothing would be taken away from Stalnaker’s formal tools, his theory would lose nothing, but gain in flexibility and power.

---

\(^{80}\) The above argument can be found in Perry, 89-91.
Stalnaker does not seem to disagree with this point, and concedes that:

“Possible worlds semantics could benefit, in some applications at least, from an account of the internal structure of the relevant propositions which permitted one to talk of the possible facts which make up the alternative possibilities” (Stalnaker 1986, 119).

But, Stalnaker maintains that if his understanding of representation is correct, then possible worlds should be taken to be complete. He argues that the:

...assumption of internal completeness is required by the explanation of propositional contents as sets of possible states of the world, and this explanation is motivated by our account of the nature of representation: since to represent the world just is to locate it in a space of alternative possibilities, content should be explained in terms of those possibilities (Stalnaker 1986, 118).

If Stalnaker were to admit partial functions to his theory, then he would have to revise his position on the completeness of possible worlds and whether or not to treat them as primitive. Since his philosophical understanding of representations and rational activity motivates these positions, Stalnaker maintains that such a change cannot be easily made. It would involve a fundamental shift in the theory because it would require some way of individuating partial worlds. This, in turn, would require recognizing parts of possible worlds as having standing in the theory, and this would mean recognizing primitives other than possible worlds themselves — most likely individuals, properties, relations, and locations.

Barwise points out that because Stalnaker is not committed to any metaphysical theory of possible worlds, there should be no motivation for him not to
accept partial functions (Barwise 1986, 431). Indeed, what Stalnaker claims is central to all rational activities is the representing of different possibilities. There really is no reason to think that such possibilities need to be represented as complete worlds. Rather, it seems more accurate to say that the aspects relevant to the situation are represented. Barwise and Perry’s theory captures, then, what Stalnaker claims is central to rational activities.

In Chapter Three, I presented Stalnaker’s theory as a step toward providing an account of meaning that could be used to describe the content of the beliefs of non-language-using animals. Stalnaker’s account, in general, works well in this capacity. Situation semantics, however, has all the strengths of possible worlds theory, and avoids some of the problems that face possible worlds theory. It seems then that situation semantics is a better option than possible worlds theory. I now turn to the question of whether or not Barwise and Perry’s work is compatible with Bennett’s. I argued in Chapter Three that Stalnaker’s work and Bennett’s complemented each other well. 81 I will argue that Barwise and Perry’s theory can play the same role, when placed in conjunction with Bennett’s, that I argued Stalnaker’s could in Chapter Three.

Like possible worlds semantics, situation semantics provides a way of representing non-linguistic content and of describing the relationships between beliefs. But, as is the case with possible worlds semantics, situation semantics does not give us an indication of how we might decide whether an organism is capable of

81 See Chapter Three, 94-102.
beliefs. This is where combining Barwise and Perry’s work with Bennett’s is particularly helpful.

The position I have adopted from Bennett is that an animal believes that $P$ if it registers that $P$ and is highly educable about many kinds of propositions, including ones like $P$. The last thing I will do in this chapter is show that Bennett’s conceptions of educability and inquisitiveness can be described in terms of situation semantics and that Bennett’s use of these to justify the attribution of beliefs is compatible with Barwise and Perry’s understanding of belief.

There are differences in the accounts of the content of mental states given by Bennett and Barwise and Perry. However, this seems to be due to the different goals of the theories rather than to contradictory answers given by the theories to fundamental questions. The primary concern in the aspects of Bennett’s work that I have discussed is in how meaningful content might be attributed to the mental states of non-language-using animals. Barwise and Perry, on the other hand, are more concerned with what constitutes meaning — for them, the systematic relationships between situations — and how mental content can be represented. This, of course, results in some differences in the way that each of the theories characterizes mental attitudes. I am not trying to integrate all of Bennett’s work in *Linguistic Behavior* completely with situation semantics. I am only claiming that Bennett’s position on the types of behavior that justify the attribution of belief is compatible with Barwise and Perry’s situation semantics.
One of the most critical distinctions considered in this thesis is Bennett’s distinction between registration and belief. The distinction between registration and belief, described by Bennett, fits roughly onto Barwise and Perry’s distinction between $R$-perception and $R$-belief. Some of the differences between Bennett’s and Barwise and Perry’s work are reflected in the fact that the categories of registration and perception are not equivalent. In Bennett’s use, saying that an animal registers information implies much less about the animal’s perceptual capacity than saying that an animal perceives something. Bennett emphasizes that very simple systems such as blades of grass and self-guided missiles register information, and explains that one reason he uses the term ‘register’ is that he wishes to avoid questions about which lower animals actually perceive. Barwise and Perry, on the other hand, seem to have the common use of ‘perception’ in mind. It is not clear whether or not blades of grass or self-guided missiles perceive anything, in this sense. They do not explicitly make any claims concerning species with low-end perceptual capacities — those organisms like grass and simple animals that are affected by information from their environments, but do not have the highly developed sensory capacities of higher mammals that allow for detailed classification of items in their surrounding environment. But they do explain that for an epistemically neutral claim to be true, the perceiver must be alert and have perceptual organs that are intact. It is difficult to see how a blade of grass, or a clam, could be said to be alert or not alert. While it is not completely clear that Barwise and Perry would, in fact, exclude these organisms

---

82 See Chapter Two, 45.
from the class of things they take to be capable of perception, it is not obvious that they would admit them either; it seems fair to say that they are not concerned with that type of question.

This does not show any inconsistency between the two accounts; it only shows that registration may be a broader category than perception — all perceptions are registrations, but not all registrations are perceptions. I will not attempt to formulate conditions for the capacity to perceive, rather than merely register, information. In the context of the issues that I am dealing with, the more important question to ask is whether or not Bennett’s characterization of the relationship between registration and belief is compatible with Barwise and Perry’s characterization of belief.

Barwise and Perry’s discussion of belief begins with a claim that they need to bring in belief to explain phenomena that cannot be accounted for by perception alone. In their discussion of Joe’s perception of Jackie biting Molly, one might want to explain Jackie’s becoming upset by saying that Molly is Joe’s dog and children become upset when they see their dog being bitten. Barwise and Perry then point out that a great deal beyond visual perception must occur in order for this to be the case: “Joe has to be able to recognize Molly as his own dog. This takes use outside of the sphere of primary perception, for being his dog is not a visual property. It also forces us to introduce ideas and concepts, so that we can discuss recognition” (Barwise and Perry 238). The claim is that Joe’s recognition of Molly as his own dog requires the introduction of concepts and ideas because they are necessary to explain things like
ownership that are not available to perception, but play a role in explaining Joe’s behavior. In order to discuss ideas and concepts, Barwise and Perry claim further that beliefs need to be added to the account (Barwise and Perry 241).

According to Barwise and Perry, perception can take place without ideas or concepts. What Barwise and Perry claim to be doing is classifying situations. R-perceptions and R-beliefs are, on their theory, indirectly classified situations. Thus, it does not necessarily follow from the fact that Joe sees that Molly is being bitten, that Joe has a concept of Molly. Instead, what Barwise and Perry would claim is that Joe is in a perceptual state that means that Molly is being bitten because of the systematic relationship between Molly being bitten and Joe’s perceptual condition. However, in the case of Joe recognizing that Molly is his dog, it is the recognition of a non-observable fact that is claimed to be motivating Joe’s behavior. So, Joe must have some concept of ownership of a pet that can play this role.

Once we know that Joe is crying because Molly is his dog and Molly is being bitten, we know that we have to attribute concepts and beliefs to Joe. Barwise and Perry explain, then, how these can be represented. But this is of little help in determining in which cases we can say that an organism is behaving the way it is because of its beliefs. Consider the case in which Molly perceives Joe being hurt by another human. There are plenty of reports of cases of dogs defending their owners in such situations. Do we then attribute beliefs to Molly? Bennett’s criteria can help in cases such as this.
Bennett's position is that if an animal can have a belief that $P$, it must be educable not only about $P$, but also about many other propositions, some like $P$, and it must be inquisitive about many kinds of propositions, not excluding any important ones which are of the same type as $P$.\footnote{If an animal was not educable about many different, but related, propositions, it would be hard to show that it had any beliefs at all. I will say nothing in support of this claim here since I discuss it in detail, and defend it, in Chapter Three. See above, 98-101.} Just as in Chapter Three I described how, in possible worlds terms, Bennett's criteria could be described, they can also be described in terms of situation semantics. This amounts only to describing educability and inquisitiveness in terms of situations since, if this can be done, it can be asked, in situation semantics, whether or not an animal is inquisitive and educable. Because this way of thinking about the behavior of animals is similar, in many ways, to possible worlds semantics, the task, at this point, is fairly straightforward. The description I give follows the general form of the description of Bennett's concepts in terms of possible worlds theory.

As Bennett describes educability, an animal is educable if it can vary behavior in order to act in ways that more often, or more efficiently, lead to its achieving its goals.\footnote{See Chapter Two, 52-54.} In terms of situation semantics, if an animal varies its behavior in ways that result more often, or more efficiently, in bringing about situations that are desirable to it, we can say the animal is educable.

If we say of an animal that it is inquisitive, we are saying that it seeks information about its environment. Bennett explains this in terms of epistemic enrichment. An animal is epistemically enriched if it finds a previously unknown
course of action open to it that will either lead to more satisfaction or will more probably lead to satisfaction than the courses of action of which the animal was previously aware. An animal may register that a particular course of action may lead to epistemic enrichment. Finally, attributing to an animal the registration that a particular course of action may lead to epistemic enrichment might explain an animal’s behavior.

Bennett’s conception of inquisitiveness can be described in situation semantics in the following way. We can say that an animal is epistemically enriched if it becomes aware of the fact that a course of action will lead to a more attractive situation, or that a course of action is more likely to bring about a desired situation than the courses of action previously known to the agent. An animal then, is inquisitive if it seeks information concerning the connection between actions available to it and situations those actions will bring about.

Unless organisms do these things, it is difficult to see why it would be necessary to attribute beliefs and concepts to them, and, if an organism is educable and inquisitive, as Bennett describes, it seems that if we want to explain its behavior, we must do so in terms of beliefs and concepts. Like Stalnaker, Barwise and Perry provide a way to represent the structure of relationships between mental states. We can, then, apply Bennett’s claims concerning how to determine which animals have beliefs to Barwise and Perry’s work. At this point, we have a theory that describes the content of beliefs and perceptions in term of situations rather than possible worlds, and we have criteria from Bennett that distinguish between things that think and
things that do not. This provides a foundation on which to base discussions of animal beliefs. In Chapter Six, I will argue that if we base our understanding of the mental content of non-language-users on this foundation, we can solve at least one significant problem in the attribution of beliefs to animals that do not use language.
Chapter Six
Non-Linguistic Theories of Content and Animal Thought

In this chapter, I will examine some of the presentations of an argument commonly given in support of the idea that language is necessary for thought. I have termed this line of reasoning the specificity argument. Proponents of it claim that if a being does not have language, the content of its mental states is not specific enough to justify the attribution of beliefs to it. The general idea behind the argument is that intentional states, such as beliefs, have propositional content and that the behavior of non-language-users could never provide the kind of evidence needed to specify a particular proposition as the content of a belief. It is further argued that the inability to attribute intentional states with specifiable propositional content to non-language-users is evidence that they have no beliefs. I will argue that while much of what is claimed in the specificity argument is true, and that there are things we can learn from it, the difficulties in attributing thoughts to animals that stem from it can be overcome by adopting a non-linguistic theory of content.

In Chapter One, I briefly mentioned the position presented by Norman Malcolm in his article “Thoughtless Brutes.”85 I will begin this chapter by presenting an exposition of, and comments on, Malcolm’s position, to which much of the literature concerning the specificity argument refers. I will then turn to versions of the specificity argument presented by Donald Davidson and Stephen Stich.86 Davidson

85 See Chapter One, 9-10.
argues explicitly that only the placement of a belief within a complex network of other beliefs will allow us to attribute content to that belief, but does not think that such a placement is possible in the case of a non-language-user. Stich, while making no explicit reference to such a network, argues that non-human animals do not possess the background knowledge on which our attribution of beliefs depends. He concludes that animals, while not believers in the full sense that human beings are, possess pseudo-beliefs or inner states that play functional roles similar to human beliefs, but which have no specifiable content.

Davidson’s claims about the holistic character of belief and Stich’s argument that beliefs depend on specific background knowledge seem to be correct. I will argue, against Davidson and Stich, that we have good reason to believe that many animals do possess networks of belief and behave in ways that allow us to attribute the background knowledge needed to meaningfully make belief attributions. My argument is based, in part, on an application of the work of Bennett, Stalnaker, and Barwise and Perry, as it was presented in the previous chapters, to issues central to the specificity argument. Stalnaker has commented on Stich’s position directly, and I will argue that his claims that non-linguistic accounts of content help with issues raised by the specificity problem are correct. However, I will show that situation semantics can provide the same remedy as could possible worlds theory to the problems that stem from specificity issues, and that it provides further advantages in

28. It should be noted that Stich has revised his position on animal thinking (See Stephen Stich, From Folk Psychology to Cognitive Science (Cambridge, MA: MIT Press, 1983) 104-6. His argument however, is still one of the clearest distillations of the difficulty in question and I make use of it for that reason.
describing the beliefs of non-language-users as well. While I have chosen to focus on Davidson and Stich, my criticisms are generalizable to other arguments of this type.

Before turning to Davidson's and Stich's arguments, I will discuss Malcolm's position in some detail. Malcolm does not present a version of the specificity argument, but his understanding of the relationship between language and thought shares many features of accounts given by proponents of that argument. I will emphasize the points that Malcolm's position has in common with these in order to make the role of the specificity argument more explicit.

In "Thoughtless Brutes," Norman Malcolm presents an example that has become central to much of the literature on the topic of ascription of content to animal thoughts:

Suppose our dog is chasing the neighbor's cat. The latter runs full tilt toward the oak tree, but suddenly swerves at the last moment and disappears up a nearby maple. The dog doesn't see this maneuver and on arriving at the oak tree he rears up on his hind feet, paws at the trunk as if trying to scale it, and barks excitedly into the branches above. We who observe this whole episode from a window say, 'He thinks the cat went up that oak tree' (Malcolm 13).

Malcolm endorses the claim made from the window but not the related claim that "The dog had the thought that p." The difference is between saying the dog thinks such and such, and saying the dog had the thought that such and such. According to Malcolm, "...although we apply the word 'think' to animals, using it as a transitive verb taking a propositional phrase as its object, we do not thereby imply that the animal formulated or thought of a proposition" (Malcolm 50). We, Malcolm argues, often apply the term "think" to humans in the same way. He gives the example of
someone who, while having a conversation, absently fumbles for his car keys in his pocket. Malcolm maintains that we would say of this person that he thinks his keys are in his pocket in the same way we say of the dog that it thinks the cat is in the oak tree — that is, without attributing a thought to him. In the case of human beings, however, we also attribute thoughts with propositional content.

Malcolm gives a similar analysis of other propositional verbs. Assume we are watching a particular person walking on an icy path in a gingerly way and say he is doing so 'because he realizes the path is slippery'. This does not imply that the proposition 'The path is slippery' crossed his mind (Malcolm 52). We could find out if the proposition had crossed his mind by asking. In fact, his telling us that that proposition had crossed his mind may seem to be the only way we could know that it had. Such questions cannot be asked of non-language-using animals:

The possession of language makes the whole difference. If a dog on a slippery path moved in an equally gingerly way, we could say with propriety that the dog is aware that the path is slippery. What further thing could we do to find out whether the thought 'This path is slippery' occurred to it, or crossed its mind? An undertaking of trying to find out whether the dog did or didn't have that thought is not anything we understand. (Malcolm 54)

But, it is not just the ability to express thoughts that is lacking. Malcolm takes a further step and argues that the lack of ability to linguistically express beliefs shows that they do not have any. He finds the idea that there are two parallel kinds of processes — the thinking of thoughts and their linguistic expressions — problematic. He argues that "This picture, if taken seriously, ought to create an uncertainty as to whether people have thoughts" (Malcolm 54). On such a picture, there is no
guarantee that the words and thoughts ever match up, in one's own case as well as in the case of others. Malcolm claims as well that, "The relationship between thought and language must be...so close that it is really senseless to conjecture that people may not have thoughts, and also really senseless to conjecture that animals may have thoughts" (Malcolm 54-5).

It is not, Malcolm explains, that he thinks that thoughts are sentences, other linguistic entities, or the utterances of such things. He argues instead, that the question of what a thought is has been poorly formed. He asks, "Why should we even be moved to ask what thoughts are, in general?" (Malcolm 55) In a given context, he claims, we can know what a particular thought is. He gives as an example an instance in which a man misses the bus that normally takes him to work: "On returning to his house he says to his wife, 'This may cause me to lose my job.' We know what his thought was. What can be said other than that his thought was that he might lose his job? What makes us suppose there might be some other kind of answer?" (Malcolm 55).

Malcolm's claim that thoughts are closely tied to their linguistic expressions, or a claim like it, is crucial to the success of the specificity argument. Without it there would only be the epistemological claim that we cannot know which specific thoughts an animal has. If Malcolm is correct, it follows from the fact that an animal does not behave in the ways that justify specific belief attributions, not only that we cannot know its beliefs, but also that that animal does not have any beliefs.
Malcolm’s distinction is between having the thought that such and such and thinking such and such. Before presenting an analysis of his position, I will attempt to clarify what he actually means by ‘thinking’ and ‘having a thought’. One might be inclined to characterize thoughts as unconscious beliefs, in a Freudian/psychological sense, but this would not capture what Malcolm is getting at. In Malcolm’s example, the dog’s thinking that the cat is in the tree is caused by a perception, just as many conscious human beliefs could be, and the role that it plays in bringing about behavior, and other thoughts, is much more like a conscious human belief than an unconscious belief of the kind discussed by Freud. It would surely be wrong to explain the thoughts that Malcolm has in mind as being repressed or playing the types of roles the Freudian unconscious beliefs play.

One might also be tempted to call Malcolm’s thoughts ‘non-occurent beliefs’. This would not be quite right either. ‘Non-occurent beliefs’ are usually taken to be beliefs that one is said to hold, but that are not, at the time in question, occurring in one’s consciousness. Consider my belief that my birthday is September 3rd. This is certainly a belief that I hold, and that I have consciously formulated in the past. But, five minutes before the time of the writing of this sentence, it was not in my consciousness and, as far as I know, it played no role in determining my behavior. Another example is my belief that the Washington monument is taller than George W. Bush — which has never entered my mind before writing this, but which I surely could be said to have believed. Non-occurent beliefs are different from Malcolm’s instances of thinking since non-occurent beliefs are either beliefs that have been
consciously formulated in the past, or are beliefs that are often the logical
consequences of other beliefs, or sets of beliefs, that are consciously formulated.
Malcolm claims that non-language-using animals do not have any beliefs that are
consciously formulated. Thus, they would not have non-occurent beliefs. As well,
non-occurent beliefs are generally not taken to explain behavior or to play a role in
causing behaviors or other beliefs. For Malcolm, the dog’s barking is explained by
the claim that it thinks the squirrel is in the tree.

Thinking and having a thought, as Malcolm uses the terms, seem to be similar,
except for the fact that the instances of thinking are not consciously formulated, while
thoughts are. For this reason, I will generally refer to what Malcolm calls ‘thinking’
as non-consciously formulated thinkings or NFT’s, in the hope of capturing the idea
that such states are supposed to play a role in directing and explaining behavior,
without actually being thought of consciously by the subject. As well, the term avoids
the problem of mischaracterizing these states as unconscious thoughts. I will
generally refer to Malcolm’s ‘thoughts’ as CFT’s or consciously formulated thoughts
to specify that he is referring to consciously formulated thoughts only.

Malcolm’s distinction between NFT’s and CFT’s seems intuitively to be
correct. There are thoughts that occur in a way the thinker is conscious of, and there
are behaviors that can be explained with reference to a being thinking such and such,
but in which he or she has had no experience of having had a thought. It might seem
to follow, and Malcolm concludes, both that animals that do not use language are
describable as having NFT’s but not as having CFT’s and that CFT’s involve
propositional content, while NFT's do not. I question Malcolm's conclusions, however, and argue that at least two issues need further consideration. First, while the distinction between cases in which thoughts are formulated and cases in which they are not can, and should, be accepted, the claim that only consciously formulated thoughts have propositional content does not follow from that distinction. Second, we should ask whether language is really the only behavior in which consciously formulated thoughts may be manifested. The answers to these questions are not obvious. I will say a few words in comment about the first. I will attempt to answer the second over the course of this chapter.

In addressing the first issue, we should consider three possibilities. There is the account of the relation between propositional content and thinking offered by Malcolm — that consciously formulated thoughts have propositional content, but NFT's do not. But, there are two other possibilities as well. It could be the case that anything we call thinking, involving CFT's or NFT's, involves propositional content — in Malcolm's terms, we might say that to think such and such involves propositional content no less than having the thought that such and such. Or, it could be the case that propositional content does not play the type of role that Malcolm takes it to in cases of either type — having thoughts or thinking. I will argue that both of the possibilities I suggest are more acceptable than the one supported by Malcolm if we take attributions of NFT's to play any serious role at all in our explanations of the behavior of animals or humans. Ultimately, however, I will argue that the third
possibility — that propositional content does not play the role that Malcolm takes it to
play, in either NFT’s or CFT’s — should be accepted over the first two.

In Chapter Three, I presented Stalnaker’s view that possible worlds, rather
than propositions, should be taken to be the principal objects of thought. In Chapter
Five, I argued, following Barwise and Perry, that we can describe mental states in
terms of situations, leaving propositions, for the most part, out of the account.
Malcolm points out difficulties in attributing propositional content to the inner states
of non-language-users that are developed further by proponents of the specificity
argument. I will argue that by adopting either of the non-linguistic accounts of
content I have discussed, these difficulties can be dealt with.

One of the problems with adopting Malcolm’s position is that, if we take
descriptions of the NFT’s of people, or animals, to provide explanations of behavior,
it is, at the very least, problematic to claim that such states have no content. It is
difficult to see how sentences such as “He thinks the path is slippery” could explain
why a person, or a dog, is walking in a gingerly way, unless they are taken to be true
statements about the content of a state of the subject in question. If we take
propositions to be the objects of CFT’s, and take the propositional content of CFT’s
to play a key role in our understanding of how attributions of CFT’s explain behavior,
then it will be very difficult to claim that NFT’s have no such content, but that claims
about a being’s NFT’s still provide explanations of behavior.

Consider what a theorist who accepts the language of thought hypothesis must
maintain if she attempts to adopt Malcolm’s position on this topic. She would have to
accept that there are sentences in mentalese for every CFT, and that there are no such sentences in the head that control behavior of which the thinker is not conscious. If there were such sentences in mentalese corresponding to NFT's, then they would have content of the same sort as the content of CFT's in mentalese, since sentences of mentalese have content. It follows from this that the strongest aspect of the language of thought theory, its ability to explain mental causation, would not apply to any NFT's. It seems likely that if the language of thought hypothesis is correct, then there are sentences of mentalese that have content and affect behavior, but never pass through the conscious mind of the believer. Thus, anyone accepting the language of thought hypothesis should reject Malcolm's claim that non-occurent thoughts do not have propositional content.

This may not seem significant to anyone who is not a language of thought theorist. However, the argument above is a sample case of a more general argument, the conclusion of which holds not only for language of thought theories, but for any theory that offers a functionalist account of mental causation. On such accounts, there is a connection between the causal properties of thoughts — what behaviors and other mental states they cause and what kinds of stimuli and other mental states cause them — and their content. According to such accounts, our semantic level explanations are taken to explain behavior because they refer to states that could also be described in terms of causal relations at the physical level. When we attribute mental states, NFT's or CFT's, we do so on the basis of behavior. If one accepts the assumption that, in correct explanations of behavior, the causal properties of mental states are products of
their physical properties, then, if a mental state is said to explain behavior, one must assume that that state is, in some sense, physically instantiated. If the attribution of an NFT explains behavior, then, according to any functionalist account, it does so only because there is a physical instantiation of the attributed state with the right kinds of connections to other states and behaviors. But if this is the case, it must have content of the same sort as a CFT. If there is no corresponding physical state, then, according to functionalist accounts, it does not explain behavior. If a language of thought theorist either must accept that NFT’s have content, or not take them to be playing a causal role in behavior, then, similarly, anyone accepting any functionalist account of beliefs is in the same position. If an NFT plays a role in such explanations of behavior, it must have causal properties. If it does, and it plays the same type of role in causing other mental states and behaviors as a CFT, then, it makes little sense to claim that such states do not have content.

The above argument is directed at those who accept functionalist accounts of mind. One can argue further, however, that on any account of the mental, if NFT’s are seen as actually playing a role in explaining behavior, then it seems unlikely that they could not have content if CFT’s have content. If NFT’s play a role similar to CFT’s and the explanation of the content of CFT’s is a result of the roles those thoughts play, then the content of NFT’s will be explainable in the same way,

---

87 It may be objected that there are many non-mental states of the body or bodily organs that do not have content themselves, but affect subjects’ mental states and behavior. For example, a heart problem certainly can cause one to be afraid that he or she will die shortly. But, NFT’s are not such states of the body. We know them through their semantic properties and attribute them in essentially the same way that we attribute CFT’s. If they explain behavior at all, according to accounts that require that mental states have physical properties, it is in ways similar to the way CFT’s explain behavior.
assuming they play similar roles. If they do not play similar roles, then there is little reason to claim that descriptions of what a person or animal is thinking explain his or her behavior.

This line of reasoning is further supported by the consideration that explanations of behavior that attribute NFT's are similar in form to those that attribute CFT's. If the claim that a person thinks the path she is walking on is slippery is supposed to explain why she is walking in a gingerly way, we must also assume that she does not want to fall and thinks that walking will prevent her from falling. Without the latter, there is no explanation. But, if it is accepted that believing (in this case a type of NFT) must be paired with desires, or desirings (as either CFT’s or NFT’s), in this way, they must have content similar to CFT’s. CFT’s and NFT’s are often attributed together in order to explain behavior. Consider the case in which I overhear a person on an icy path say to herself, “The path looks slippery” just before beginning to walk in a gingerly way. It is not obvious that she also had the CFT that she did not want to fall. If the question is asked of her or anyone else, why noticing the path was slippery would prompt her to walk that way, a good answer to it would be, because she did not want to fall and believes that walking that way will improve her chances of not falling. We would not then also conclude that this reasoning was explicit or occurred to her in any conscious way. But if we can mix NFT’s and CFT’s in this way, on the basis of semantic or intentional content, and claim that they really do influence each other and behavior, it seems that they both must have such content.
This is barring the possibility, of course, that NFT's play a role of a completely different sort than CFT's. It seems unlikely, however, that we will develop a successful theory of mental content and that there will be states that we still call mental states, and that we claim play a role in explaining and determining a subject’s behavior and other mental states in the way CFT's do, but do not have content in the same way that CFT's do. If one accepts that CFT's have propositional content, then one must either accept that NFT's also have propositional content, or see NFT's as not playing an actual role in explaining and determining behavior.

Malcolm approaches the subject from a very different perspective than the one I have taken. He does not ask questions about the causal and explanatory roles of thinking or thoughts. Instead, his primary concern is with how terms are used in ordinary language and with developing a consistent usage of the terms ‘thinking’ and ‘thought’. Still, he draws important conclusions, with significant implications, about the role of language and the limits of the abilities of animals. I am, of course, arguing that his conclusion is mistaken. In particular, I am claiming that he puts far too much emphasis on the experiences of thoughts as sentences in one’s consciousness. My suspicion is that this results from the fact that, for Malcolm, the only way that propositional content can be represented is linguistically, and the only possibility of that, for him, is in the speaking, or conscious experience, of a public language.

I will not pursue this further and I will not attempt to tackle the question of how it might be that non-occurrent thoughts have content or cause other mental states and behavior. Issues related to, and the implications of, the distinctions between
CFT's and NFT's are important. But, they are outside of the scope of my project and the attempt to pursue them would lead far afield from the aspect of Malcolm's position I wish to pursue in detail: the claim that non-language using animals do not have beliefs with content.

I will argue that it is possible for non-language-using animals to have contentful beliefs. My argument depends on the acceptance of the third option I stated above — that propositions do not play the role in NFT’s or CFT’s that is attributed to them by those who take them to be the objects of thought. As I have argued, it is not necessary to understand propositional content as being essentially linguistic in structure. This is particularly important to the issues I discuss in this chapter because, as I will explain, if propositional content is understood to be linguistic, unless one adopts the language of thought hypothesis, Malcolm’s position is almost unavoidable. Stalnaker’s work on a non-linguistic account of propositional content, as described in Chapter Three, is particularly relevant to this point, but, ultimately, I will rely on situation semantics to solve the problem.

If propositions are taken to be linguistic, there are, very broadly speaking, two ways in which thoughts could be said to have propositional content. The first of these depends on the language of thought hypothesis, the second on the possession and use of a public language. On either theory, any animal said to have beliefs, since beliefs have propositional content, must be said to have the ability to either represent propositions or to, at least, be in a state that can be linked to propositions in some way. Neither of these approaches requires that propositions are actually stored in the
brain. It might be argued, for example, that propositions are linguistic entities that exist independently of any particular being’s thinking of them, but that they can be represented in language. In the case of animals representing linguistically structured propositions, if we posit the existence of a language of thought, the issue of how animals could represent the content of a proposition seems to be less of an issue. Sentences in mentalese have linguistic structure and through their structure, represent linguistically structured propositions. This propositional content can be described in English; this is, according to such a theory, what we humans are doing whenever we describe our thoughts in public language. Adopting this theory would solve many problems, but, as I point out in Chapter One, and argue, following Stalnaker, in Chapter Three, there are several very good reasons to reject the language of thought hypothesis.

On the other hand, if propositions are taken to be linguistic in nature, and the language of thought hypothesis is rejected, we are left with the question of how animals might grasp them at all. In the case of humans, it can be argued that the use of language allows one to formulate concepts having structure and content of a sort similar to propositions. Animals, however, without language, are seen to be lacking in this capacity.

If propositions are linguistic, it is difficult to answer the question of how an animal that does not use a public language and does not have an internal language of thought could be in an intentional state that is easily and accurately described in language. It is certainly convenient to argue, if one thinks that propositions are
independently existing linguistic entities, that the grasping of a proposition, or having a belief, requires that one have the ability to use, or at least understand, a language. Arguments of this type, like the arguments that assume a language of thought, lose much of their force if propositions are not taken to be linguistic in nature.

Malcolm makes an important point in claiming that we should not assume that there are two parallel processes, thinking thoughts and expressing thoughts in language. The worry that he refers to has to do with his concern that if we understand thought as being a process that is then expressed by a description in language, then the two might never match up. I do not dispute that accepting the idea that there could be a stream of states with well-defined intentional content, which then gets described, opens the door to this type of problem. But, Malcolm’s stated conclusion that using language is necessarily tied to having thoughts should not be accepted without question. This is because he describes the difficulty as being narrower than it actually is and, consequently, his solution is also narrower than it should be.

The problem that Malcolm refers to arises in any case in which it is claimed that there could be a stream of states that have intentional content but are completely independent of the subject of those states’ interaction with the environment. In such cases, we never really know if our linguistic descriptions truly reflect the content of those states. This is not just because there is a gap between language and thought according to this understanding of the mental. It is because there is a gap between thought and any observable phenomena. Malcolm’s way of closing this gap is to claim that all thought content is related to linguistic behavior. This is only justified,
however, if propositional content could not be manifested in other types of behavior. Thus, the lesson that we should take is that the content of one's thoughts is intimately tied to one's behavior.\textsuperscript{88} It is a further question as to whether only language users behave in the kinds of ways necessary for one to have contentful mental states.

The specificity argument is an attempt to answer this question. It can be understood as supplementing Malcolm's position by attempting to show that, in fact, only creatures that use language behave in ways that are needed to have mental states with content that is specific enough to support a system of beliefs. I will address the specificity argument shortly. What Malcolm makes prominent is that if a creature does not behave in ways that a creature with beliefs would, it makes little sense to assume that it has beliefs. But, if one does not accept the claim that propositional content is linguistic in structure, then one can avoid Malcolm's conclusions that only language users have thoughts. After discussing Davidson's and Stich's treatment of Malcolm's example and others that are similar, I will return to this issue.

Davidson maintains of thoughts, in general, that it is only possible to identify and individuate them if they are located within a complex network of interrelated beliefs. Applying this claim to Malcolm's dog, Davidson says:

\begin{quote}
If we really can intelligibly ascribe single beliefs to a dog, we must be able to imagine how we would decide whether the dog has many other beliefs of the kind necessary for making sense of the first. It seems to me that no matter where we start, we very soon come to beliefs such
\end{quote}

\textsuperscript{88} I am certainly not advocating the adoption of a behaviorist position. It is not that mental content is identical with behavior patterns or dispositions to behave in certain ways. I am assuming the truth of two claims, but neither is particularly controversial. First, that content is not attributable unless behavior is observed (spoken or otherwise) and second that in the absence of any interaction with an environment, there could be no mental states that have content.
that we have no idea at all whether the dog has them... (Davidson 1985, 475).

In order for the dog to have a belief about a tree it must have many true beliefs about trees in general and about the particular tree in question. Although there could not be a specific list of beliefs one must have in order to have beliefs about trees, there must be enough related beliefs to make the attribution of beliefs about trees sensible. How could one have a belief about a tree if he or she had no beliefs about the general nature of trees — that trees are plants, that trees are made of wood, or that trees have leaves? If we are going to make sense of a particular belief, we must situate it within a group of beliefs that are related to it. In order to do this in the case of the squirrel-chasing dog, we need to establish which specific related beliefs the dog has and, according to Davidson, this is not possible without building a complex network of beliefs. For this, he claims, language is required.

If we accept that the dog believes that ‘The cat went up the tree’, Davidson asks, how can we tell which related propositions the dog may also be said to believe? If we ask questions such as ‘Does the dog think that the cat went up the oldest tree in sight?’ Or that ‘the cat went up the same tree it went up last time the dog chased it?’ Davidson claims that, “It is hard to make sense of the questions” (Davidson 1985, 474). If we cannot answer such questions, then we cannot answer any specific questions about what exactly it is that the dog believes. If we cannot answer specific questions about what the dog believes, argues Davidson, it makes little sense to speak of the dog believing anything.
Stephen Stich also discusses the possibility of attributing beliefs to a dog and constructs an argument that is similar in many ways to Davidson’s, but which has a conclusion more explicitly in line with Malcolm’s. We are asked to consider whether or not the dog Fido can have thoughts about bones. Stich claims that although it is true that Fido might be able to recognize bones as a category of objects (that is, Fido treats all bones in a roughly similar manner) and further that Fido has treated an object in the yard as a bone, we still cannot claim that Fido has the belief that ‘There is a bone in the yard.’

Fido does not, according to Stich, have the right kind of knowledge to justify the attribution of beliefs about bones to him. He does not, for instance, have beliefs about such things as the origin and general anatomical function of bones. Nor would Fido recognize atypical bones (such as the bones of the inner ear or the collar bone of a blue whale). Finally, Fido could not tell the difference between real and fake bones, (plastic or other doggy treat type bones). It is not just that Fido might make a few mistakes, the same is true of many humans. More importantly, in the case of an atypical bone, there is no way for Fido to conceptualize the fact that the collarbone of a whale, for instance, is a bone or that a fake bone is not a bone. The distinction between real and fake could not even be made to make sense to Fido. Stich concludes that:

...given Fido’s conceptual and cognitive poverty in matters concerned with bones, it is surely wrong to ascribe to him any belief about a bone. To clinch the point, we need only to reflect that we would certainly balk if the same belief were attributed to a human who was as irremediably ignorant about bones as we take Fido to be (Stich 19).
Stich's point is not that we cannot know what categories the dog uses, or that it makes no sense to ask certain questions about what it believes, but rather that we can know that it does not have certain types of beliefs that are crucial to having bone beliefs. Obviously, the argument can be repeated for any concept in question.

According to Stich, beliefs have two aspects: they play a psychological or functional role (defined in terms of their relation to, and interaction with, other inner states, behavior and perception), and they have a propositional content. Stich argues that there may be entities who, while having states which play the functional role that our beliefs do for us, have an inner organization of beliefs which is so different from our own that we cannot specify the content of those beliefs. When we attribute beliefs to other humans of our own culture and time, both the functional role and the content of our beliefs match up. However, Stich argues that: "trouble looms when we attempt to apply the notion of belief to more exotic subjects, be they human, animal, or artifact. It is here that the two features of our ordinary conception of belief pull apart" (Stich 1979, 26). What we are left with on this view, is a kind of pseudo belief, "...a belief-like state which lacks a specifiable content..." (Stich, 1979, 27). Stich's claim is that this explains why intentional descriptions work when it comes to explaining the behavior of many animals, even though there is no specifiable content to those beliefs.

This position then, is similar to Malcolm's. Stich attempts to provide an explanation of why beliefs with non-specifiable content can play a role. But, if such beliefs play a functional role similar to the ones played by human beliefs, they cause
behaviors and mental states similar to the ones caused in humans. If this is the case, it is unclear what it would mean for their content to differ. If an animal or human from a culture very different from our own had beliefs with content that was significantly different from any belief in our culture, it is not clear why this would not be reflected in behavior in a way that would allow us to, with proper techniques of examination, discover and adjust our belief attribution. In the case of animals, again, if beliefs play functional roles similar to the roles that beliefs play in humans, unless we have reason to believe that their content is different from the content of human beliefs, we should assume the content is the same. Of course, we do have reasons to think that animal beliefs may have content that is very different from the content of human beliefs; animals behave in very different ways than humans. The question is whether these differences in behavior prevent animals from having beliefs. This brings us back to the specificity argument. Like Malcolm's position, this part of Stich's position needs the support of the specificity argument. I will return to that topic.

The arguments of both Davidson and Stich rest on the idea that we cannot attribute certain fine-grained beliefs to non-language-users. For Davidson, we cannot attribute beliefs with the accuracy required to build a network of beliefs with enough interconnections to hold itself up, so to speak. For Stich, animals simply do not have the ability to have certain beliefs that are fundamental for other beliefs. In both cases, the problem is essentially the same: even very simple beliefs depend necessarily on complex combinations of others, and there is reason to think those others cannot be held by non-language-using animals. If these claims are correct then it seems that we
do not, and could not, accurately attribute beliefs to animals. While Malcolm simply states that we do not attribute thoughts to animals, Davidson and Stich provide analyses of why it seems that we cannot do so. Davidson takes the conclusion a step further than Malcolm and Stich, and argues that if we cannot speak of the content of thoughts, it must be incorrect to speak of what a non-language-using animal is thinking. If my arguments are correct, however, the differences in the strength of the conclusions is irrelevant, since I will argue that content is attributable to animal beliefs.

One response to arguments relating to specificity is to claim that our inability to attribute specific beliefs is not as much of a problem as it might seem. K.V. Wilkes offers this view, stating:

That we find it difficult to specify the thoughts or beliefs of the dog that chases a cat up a tree, and continues barking excitedly up it even when the cat has escaped to another tree — this fact should not...attract nearly as much alarm and quibbling discussion as it has so far (Wilkes 179).

Wilkes’ reasoning is that we deal with the same type of indeterminacy when attributing concepts to adult humans on a regular basis. As well, she argues that, in terms of the concept of bone, “...the dog is not much worse off than the average adult human who has never been told of the differences between vertebrates and invertebrates; nor the pre-school child” (Wilkes 179). She claims the greater the extent to which we share broad networks of interconnected beliefs with those with whom we participate in discourse, the more straightforwardly we can attribute

---

specific content to their beliefs. But, Wilkes emphasizes that there is a lot of latitude in the amount of specificity required to attribute a belief and that our ability to do so is always measured by degree. She claims that, in the case of many human beliefs, there is no right answer to questions about what it is, specifically, that someone believes. She concludes as follows:

I am thus untroubled by the difficulty of specifying (with our linguistic categories) the contents of a cat’s, rat’s or bat’s mind. The difficulty is no worse, or better, than that of pinning down the contents of the thoughts and anticipations of a composer or an artist, or of a non-scientist about abstruse or difficult theoretical entities. There will be a huge spectrum between vagueness and precision amongst our ascriptions of what is said to be thought, believed or felt (Wilkes 180).

I do not dispute Wilkes’ assessment of the lack of specificity in many of our attributions of beliefs to humans. I support a position similar to hers on this point, in the claims I make against the specificity argument later in this chapter.90 As well, in general, I agree with her claim that we should not be troubled by the lack of specificity in our attributions of beliefs to animals. But the perspective from which I am considering the beliefs of animals is different from that which we normally attribute beliefs to humans. I am considering questions about whether or not animals have beliefs at all. While we might not be worried about a lack of specificity in attributions of beliefs for everyday interactions and interpretations of animal behavior, in addressing arguments to the effect that animals do not have beliefs, it will be helpful to be able to specify, as accurately as possible, the content of the mental states of animals. I am arguing that we can better represent the beliefs of animals by adopting a non-linguistic account of content. In some cases, especially

90 See below, 218-21.
when dealing with arguments concerning whether there is so much indeterminacy in
the beliefs of animals that is does not make sense to attribute any beliefs to them, this
will be helpful.

Properly addressing the arguments Davidson and Stich present requires being
as clear as possible about how the content of the beliefs attributed to animals can be
represented and described. One of the reasons the specificity argument has seemed so
convincing is that it is difficult to precisely describe the content of animal beliefs. At
least part of this difficulty stems from the fact that, when we attempt to describe the
belief of a non-language-using animal in language, we can’t find sentences that match
up very well with any belief that it would be reasonable to attribute to the animal.

This should come as no surprise. For the non-language-using animals we
regularly interact with, we can expect that any attempt to describe their beliefs in
language will be somewhat distorted. We see similar problems when we try to
describe the beliefs of humans who do not speak the same language as us or are from
cultures different from our own. Often, translation from one language to another is
neither simple nor straightforward. It does not follow from this claim that there are no
concepts to translate.

Of course, animals do not use language, so the analogy is not perfect. Rather
than using this point to argue in favor of the idea that non-language-users have
thoughts, I want to emphasize that if they do, it is a difficult question as to how their
content can be described. The problem with the content of animal thoughts may not
be that there is no way to specify their content as much as it is that there may not be
sentences in human language that have the same content as the thoughts of non-language-using animals. As in some cases when one is trying to express, in English, the meaning of a sentence of another, very different, language, it is better to describe what is being said than to attempt to find an English sentence that is equivalent to the sentence in the other language. Non-linguistic accounts of content provide a systematic way of doing this.

Stalnaker provides a brief argument to the effect that the application of his work to Stich’s will help solve the problem Stich raises in attributing beliefs to animals.91 Stalnaker claims that the argument Stich presents: “...is difficult to answer if we regard propositions as linguistic items or as quasi-linguistic complexes that are stored in the mind or the brain” (Stalnaker 1984, 63). As Stalnaker describes the situation, if we accept that all content is linguistic, considering Fido, the problem is: “to translate as accurately as possible the dog’s mental language into English, or to find English expressions for the concepts that are he constituents of the content of its propositional attitudes” (Stalnaker 1984, 63). If propositions are thought of in this way, he argues, it is plausible to conclude that animals do not think at all, only because it is implausible that animals have any conceptual or linguistic structures of this type.

The sets of beliefs an animal may have may be far less sophisticated than those of humans. In addition, through language, we have the capacity to easily make very fine distinctions concerning different possibilities. Still, Stalnaker maintains that

91 See specifically, Stalnaker 1984, 63-5.
we can use language to describe belief states that may not be as complex as human beliefs, and may be based on different concepts:

...our language, with its complex structure and ability to make more subtle discriminations and describe more distant possibilities, may be used to distinguish between the few alternative possibilities represented by the dog without thereby attributing our concepts to it. Because propositions do not mirror the sentences that express them, it is possible to use sophisticated, semantically complex sentences to ascribe attitudes to creatures with very limited cognitive capacities (Stalnaker 1984, 63).

The idea here is that even though the English language is based on certain conceptions of types and set up to make fine discriminations, it can still be used to describe many animal beliefs. According to Stalnaker, accurately describing an animal’s intentional states is accomplished by describing sets of possible worlds. The sentences used to describe such sets of possible worlds are not themselves equivalent to the sets they describe. Thus, we do not face the difficulty of trying to find sentences that are English translations of Fido’s beliefs. It is Stalnaker’s position that recognizing this, and adopting a possible worlds account of content, solves the difficulties presented through the specificity argument. Stalnaker addresses Stich’s version of the argument directly, but his arguments apply to Davidson’s version of the argument as well.

After presenting Stalnaker’s discussion of Stich, I will describe how situation semantics can be applied to the same issues. If animals do have beliefs, we can improve the accuracy of our descriptions of the content of those beliefs by improving the systems in which we represent the beliefs of animals. I have argued that Barwise and Perry’s theory is superior to Stalnaker’s in important ways. Because Stalnaker
has done work in addressing the specificity argument, I will review his claims first, and then turn to a discussion of how situation semantics, as described by Barwise and Perry, can not only accomplish the same work, but in fact is better able to account for our ability to express animals beliefs than is Stalnaker's.

Stalnaker discusses the issue of how, on his account, the differences between animal and human beliefs can be dealt with. Through this issue, he approaches the specificity argument. Because of the differences in background beliefs about bones between Fido and almost any human, Stich argues that it is not possible for Fido's beliefs to have content that is meaningful. In contrast, Stalnaker argues that Fido and his master can share beliefs.

Stalnaker replies to Stich as follows. When Fido's master believes there is a bone in the yard, that belief implies that what is in the yard is a real bone, not a fake bone of the type Stich mentions. Fido's belief does not imply this, since, as Stalnaker puts it: "he does not distinguish the actual situation from a similar counterfactual situation with an ersatz bone substituted for a real one" (Stalnaker 1984, 65). Still, according to Stalnaker, there is a sense in which the dog and master believe the same thing. Stalnaker considers Fido's belief in relation to the dog's master's:

They both divide the set of alternative possible ways things might have been that is relevant to explaining Fido's capacities in exactly the same way. Since propositions are functions, the identity conditions are relative to a given domain of arguments. Two expressions may determine the same proposition relative to one set of possible states of the world, but distinct propositions relative to larger set (Stalnaker 1984, 65).
Stalnaker's claim that the dog and master divide the set of alternative possibilities in exactly the same way is interesting. Given that Fido is not able to distinguish ersatz bones from real bones it would be seem that the set of possible worlds that would correspond to Fido's belief that the bone is in the yard would differ from Fido's master's belief that there is a bone in the yard. Even when comparing the beliefs of two humans, we expect some differences, based on differing experience and background beliefs about their conceptions of objects such as bones. It would seem then, that it would be rare to ever find two people who divide the space of possibilities, on any subject, in exactly the same way.

Stalnaker's way of dealing with this issue is to focus on the fact that questions about beliefs are relative to contexts. He argues that the question one is asking determines the set of possible worlds to be considered in the individuation of beliefs. Beliefs, according to Stalnaker, are described in terms of goals and choices concerning which actions will lead to the actualization of a possible world. In this case, the things that could be involved are Fido, Fido's master, the house, the back yard, and the bone. Fido's possible reactions to the bone being in the yard are fairly limited. Fido can attempt to get outside and dig up the bone, then either chew on it, leave it, or rebury it, or, Fido can ignore the bone's existence. Each of these actions will give us evidence as to what Fido's goals and beliefs are. The possible states of affairs might include one in which Fido is released into the yard and chews on the bone, one in which Fido is released into the yard, but does not chew on the bone, and one in which Fido is not released into the yard at all. What Fido must do is determine
which action will lead to the state of affairs he wants to bring about. In this case, the
domain of arguments — the possible worlds relevant to Fido’s choice — is fairly
limited. The key to the claim that Fido and his master have the same belief is that
Fido’s master has the same relevant options to consider, and thus the two are working
with the same domain of arguments. While each may describe the possible worlds in
question differently, they are both choosing behaviors on the basis of the same set of
relevant options. This is the sense in which both Fido and Fido’s master have the
same belief.

Relative to a larger set of possible worlds, the two may be said to have
different beliefs. A slightly larger set might include not only one bone but two, one a
real bone and one fake bone. In such a case, Fido might treat both bones in the same
way under all conditions, while Fido’s master might, at times, treat the fake bone
differently — for example, the master might allow Fido to bring soy bones into the
house, but not real bones.92 If Fido treats fake bones and real ones the same way, and
Fido’s master does not, then it seems clear that relative to the set of possible worlds in
question, the two have different beliefs.

Stalnaker points out that one could argue that two beliefs are not really
identical unless the propositions determined by the beliefs have the same value in all
possible worlds. He responds to this possible claim with a reminder that his account
of possible worlds does not even claim that such a set exists (Stalnaker 1984, 64).

Stalnaker maintains that his account of possible worlds is not a metaphysical account,

92 In such a case, it is not unlikely that Fido could learn to distinguish between soy bones and real
bones fairly quickly. However, the point that I am making should be clear.

200
but instead is an account of the structure of logical activities. His position is that possible worlds are what are differentiated in rational activities and that his theory of possible worlds describes the logical structure and relationships between intentional states. In order to do this, one does not have to posit the existence, independent of human activity, of the set of all possible worlds. Even if one does posit such a set, it does not follow that all members of the set are relevant to all rational decisions. If the options are as limited as they are in the case described above, which involved only Fido, a real bone, the yard, and Fido’s master, there is no reason to consider any options beyond those options that are choices for Fido and Fido’s master. Since, in this case, no decision involves a distinction between a real bone and an ersatz bone, this distinction is not relevant to Fido’s, or Fido’s master’s, belief.

As the case mentioned above concerning my beliefs about medical procedures and my doctor’s belief shows, in the case of human beings, we certainly do not, in regular usage at least, require that two people have exactly the same level of sophistication or indeterminacy in their mental states to say that they have the same belief. If we truly accepted the claim that for two beings to have the same belief, the belief of each being must imply the same truth value for exactly the same set of possible worlds when all possible worlds are considered, we would have to radically change our practices of belief attribution for human beings.

Even with these considerations, however, it is important to note that our having the same beliefs as animals is not an essential component of the claim that animals have beliefs or that humans and animals share any beliefs.
at all. It is essential to our attributing beliefs to animals, however, that we have a system in which we can accurately represent animal beliefs. The strength of the specificity argument rests on the idea that there is no way that we can, with any accuracy, describe the content of beliefs of non-language-using animals with enough specificity to make sense. Stalnaker’s theory allows us to attribute content in a way that preserves the logical structure of belief systems, but does not depend on the believer using a language. This allows one to show how beliefs may be related to each other and interconnected.

Stalnaker provides a plausible answer to the question of when different subjects can be said to share beliefs, but situation semantics offers a better account of the similarities and differences between the beliefs of dogs and their masters, and animals and humans in general. I explained in Chapters Four and Five how taking possible worlds to be complete and primitive leads to problems stemming from the inability to distinguish between logically equivalent statements with different content.

That Stalnaker conceives of possible worlds as complete and primitive also limits the kind of answer he can give to the question of whether Fido and his master can share beliefs. Stalnaker claims that when Fido and his master have beliefs about a bone in the yard, they are considering issues about ways the world might be in the present and in the future. The set of possible worlds that should be considered as representing the content of either of their beliefs,
according to Stalnaker, includes only those relevant to the ways the world might be, given the objects present — the dog, master, bone, and back yard — and the possibilities they yield, now and in the future. The sense in which the dog and master share beliefs depends on the claim that it is not part of the content of the master's belief that the bone in the yard is not a fake bone of some type, but an actual bone. In this case, the master's further knowledge does not bring it about that there are, in fact, more possible ways the world could be, given the issues under consideration and the objects present. Making this move provides a solution to the problem, but it is not clear that it is the best solution that could be offered.

In particular, it is not clear why we should accept that the content of Fido's master's beliefs should be determined according to the set of possible worlds that he and Fido share, rather than on its own terms. While Stalnaker is right to say that the domain of arguments should be limited in terms of what is being determined, there still should be some way of accounting for the differing concepts of the dog and master. It seems that, at least in the majority of cases in which they interact, Fido and his master will not consider equivalent sets of possibilities.

The most accurate way to describe the relationship between Fido's beliefs and Fido's master's would be one that captured the fact that they are about the same thing, but that the dog and human have different understandings of the objects involved. Stalnaker's theoretical apparatus does
not allow him to describe this type of relationship. Stalnaker can say things such as, the set of possible worlds that are true according to Fido's master's belief is equivalent to, or a subset of, the set of worlds that are true according to Fido's belief. He could say that, since Fido does not distinguish between real and fake bones, and Fido believes there is a bone in the yard while his master believes that there is a real bone in the yard, the set of possible worlds in which the belief is true for the master is smaller than the set of possible worlds in which Fido's belief is true. But, this tells us little about the relationship between the objects in the sets. According to possible worlds theory, content is fully determined on the basis of sets of complete worlds. So, the set of possible worlds in which the master's belief that 'there is a bone buried under the tree in the yard' is true, is a subset of the set of worlds in which the master's belief that 'there is a bone in the yard' is true. This does not explain how and why the sets are, or are not, related.

We could much more accurately represent the situation if we could describe the ways that the content of one belief overlaps with the content of another belief, or describe what is shared and what is not when two believers have the same belief but one has a more complete understanding of the concepts or situation involved. Situation semantics can tell us much more about the ways that beliefs are related to other beliefs. As explained in Chapter Five, John Perry argues, in “From Worlds to Situations” that Stalnaker would do well to extend his theory to include partial functions and partial worlds. He further argues that these additions would make
Stalnaker’s theory a version of situation semantics. If Stalnaker expanded his theory in order to include partial ways, his theory would not be built on possible worlds, but on parts of worlds or situations. One possibility here would be to extend Stalnaker’s theory in this way. I am not going to pursue the possible development of Stalnaker’s theory as Perry suggests. Instead, I will turn to a brief discussion of how situation semantics applies to examples such as those involving Fido and his master.

Perry refers to ‘ways’, or ‘total ways’, meaning what Stalnaker does when he uses the term ‘possible worlds’. Ways provide answers to every issue under consideration. Partial ways are parts of total ways. Perry’s suggestion is that Stalnaker’s theory could be extended to include partial ways in addition to total ways. A theory that allowed for partial ways in addition to total ways would allow one to represent the way in which the content of Fido’s belief is a part of the content of Fido’s master’s. As Perry explains, in reference to Stich’s example:

Suppose the master has the concept of an ersatz bone, but does not have an opinion as to whether the buried bone is ersatz or not. Then we can say that the master believes,

that the bone is buried and it is ersatz or it is not ersatz

while the dog does not believe this, but only,

that the bone is buried (Perry 89).

Perry argues further that:

It seems that we might find it useful to be able to make the distinction between these two propositions, within one set of alternative possibilities, in giving semantics for something like:

---

93 See Chapter Five, 160-2, for a more complete exposition of Perry’s argument.
The dog believes the bone is buried, while the master believes that and, in addition, that it is ersatz or not ersatz (Perry 89-90).

Perry’s position in “From Worlds to Situations” is that Stalnaker should expand his theory to incorporate partial ways and propositions that provide partial functions from ways to truth-values. It is worth pointing out that, if situation semantics can represent this difference, there is no reason it could not also represent any overlap or dissimilarity between one of the master’s beliefs and the one of the dog’s beliefs.

Given the situation Perry describes above, there are two options left open to situation semantics to explain the relationship between Fido’s belief and his master’s. I will outline both. In the first, we will approach the situation in a way similar to Stalnaker, according to which, Fido and Fido’s master share the same belief. Then, I will discuss a more complicated situation in which Fido believes that there is a bone in the yard and Fido’s master believes that there is a bone in the yard and that it is not ersatz.

In the former case, we can say of both Fido and Fido’s master that they have the belief that there is a bone in the yard. In Barwise and Perry’s system, this can be represented:

\[ e_2: \text{at } t: B, \ Fido, \ E; \text{yes.} \]
\[ \text{of, } a', \text{ bone; yes} \]
\[ \text{of, } b', \text{ yard; yes} \]

where
\[ E: = \text{at } h: \text{buried in, } a', b'; \text{ yes.} \]
\[ \text{bone, } a'; \text{ yes} \]
\[ \text{yard, } b'; \text{ yes} \]

As well, Fido’s master’s belief can be represented in the same way:
in $e_0$: at $l$: $B$, master, $E$; yes.
of, $a'$, bone; yes
of, $b'$, yard; yes

where
$E$: at $h$: buried in, $a'$, $b'$; yes.
   bone, $a'$; yes
   yard, $b'$; yes

As to questions concerning how both Fido and Fido’s master can have the same beliefs when they have very different concepts, both of which we refer to with the term ‘bone,’ Barwise and Perry could give an answer that is similar to Stalnaker’s, but given in terms of situations rather than worlds. For many purposes, in cases in which the principal concern is with how the bone will be treated and in which the situation is roughly as described by Stalnaker, this description suffices. Both Fido and his master are considering the same set of objects and possibilities. Given that both have an inner state that is about the bone, and the backyard, what we are concerned with, and what Fido’s master is concerned with if he tries to explain why Fido is looking intently at the door that leads to the yard, are the same.

Stalnaker claims that both Fido’s and his master’s concept would assign truth values to the relevant set of possible worlds in the same way, and thus, the two can be said to share a belief. Situation semantics can give a similar answer. The concepts of both the dog and the master are anchored to the world with the same objects and the situation each believes is the case is the same. The possible behaviors available to the dog and the master can be understood in terms of each recognizing the situation, or state of affairs in
question. In many cases, this approach will be adequate to explain situations in which the dog and master interact.

But, Barwise and Perry's theory provides another possible way of explanation. Since their theory allows us to see some situations as parts of other situations, we can see Fido's belief as part of his master's. Another interesting case is one in which Fido has the belief described above and Fido's master has the same belief, except that the master believes that the buried bone is not an ersatz bone, or that it is the bone of a cow or anything else different from what Fido's concept allows. Such a belief could be represented as follows:

\[
\text{in } e_o: \text{ at } l: B, \text{ Master, } E; \text{ yes.} \\
\text{of, } a', \text{ bone; yes} \\
\text{of, } b', \text{ back yard; yes} \\
\text{where} \\
E: = \text{ at } h: \text{ buried in, } a', b'; \text{ yes.} \\
\text{bone, } a'; \text{ yes} \\
\text{not an ersatz bone, } a'; \text{ yes} \\
\text{back yard, } b'; \text{ yes}
\]

This gives us a more accurate characterization of the relationship between the dog's belief and the master's in cases in which it is relevant. In this case, they may not be said to share a belief, but it is easily seen how they might communicate or predict each other's behaviors. The dog's incomplete understanding of the master's state may not necessarily interfere with its reacting appropriately or predicting the master's behavior. Similarly, the fact that the master's belief includes more information than the dog's does not at all limit the master in attributing mental states to the dog.
This case is in many ways similar to ones mentioned above, in which human patients communicate with human doctors or any other experts. It is awkward to say that I have the same concept of bone that the average orthopedic surgeon does, but it would be false to say that we cannot communicate about bones. The fact that situation semantics allows this flexibility in approach to differences in specificity of concepts is another way in which it is a more workable theory than Stalnaker’s in the area of explanation and representation of animal beliefs.

We can consider the case in which Fido and his master have very different concepts of bone as well. As Stich points out, Fido likely does not have any understanding of the anatomical function of bones and may not be able to distinguish between bones and other chewable things. It may be then that Fido’s concept of bone has little in common with Fido’s master’s. A strength of Barwise and Perry’s theory is that it allows us to be more clear about what Fido and his master share and what they do not.

Fido and his master might have dissimilar concepts of bone, but, in this case, each of their concepts is anchored to the same objects. If we apply Barwise and Perry’s explanation of indirect reference, we can refer to Fido’s inner state because there is a uniformity between Fido’s inner situation and the situation in the back yard.

Fido’s master and Fido, then, might not share a view of the situation type, but might still be able to interact with each other and predict each other’s
behavior with accuracy. At the very least, since both concepts of bone are anchored in the same objects, there is likely to be some overlap in the qualities Fido and his master attribute to the object in the back yard. Again, the case might be similar to one to one involving my doctor and me. Or, it might be similar to one involving a professor and students. It is possible that some first year philosophy students and their professors have concepts of philosophy that are nearly as different as Fido and Fido’s masters concepts of bone. Still, students and professors in such situations often get along fairly well, have conversations about the nature of philosophy, and work together. Adopting Barwise and Perry’s theory helps to provide a picture of how we might characterize and compare the beliefs of humans and animals, and answers questions about how it is that beings with very different concepts and beliefs interact successfully.

What I have argued so far supports the claim that if non-language-users have mental states with content, we can represent that content using Barwise and Perry’s situation semantics. What are eliminated by the adoption of a non-linguistic theory of content are problems relating to representing the content of animal beliefs. This does not, however, show that the specificity argument is wrong. I am now in a position to address the specificity argument more directly. Before turning to arguments that I think provide good reasons to reject the conclusions of Davidson, Stich, and Malcolm, I will examine one response, that of David Armstrong, that attempts to accomplish this, but is, in
my mind, unsuccessful. I do this in order to make the difficulty that Davidson
and Stich refer to more explicit. In addition, I will incorporate part of
Armstrong’s work into the solution that I offer to the problems raised by
observations related to the specificity argument.

Armstrong’s reply to the specificity argument is that since the problem
is that we cannot say exactly what animals’ thoughts are, we might avoid
attempting to attribute thoughts with precise intentional content until we have
gathered enough evidence to attribute content to beliefs with authority.94 He
argues that we could even phrase our statements about animal beliefs in such a
way that they do not depend on our knowing the specific content of the beliefs
which are referred to. To use Davidson’s example, one might say, “The dog
thinks, with respect to that oak tree, that the cat went up it” (Davidson 1985,
475). This allows us to speak of animal beliefs while recognizing that we are
not capturing the precise content of those beliefs.

This solution assumes that every belief has some specific content; it is just
that we do not, at the moment, have access to it. But, according to Armstrong, our
ignorance need only be temporary. We could rely on careful study of animal behavior
to come up with exactly what the contents of those beliefs are. Research could
uncover exactly how Fido breaks down the world and thus exactly what the contents
of Fido’s concepts are.

Unfortunately, Armstrong’s solution does not address the real concern. The problem according to Davidson and Stich is not just that we cannot know the beliefs of non-speakers. They make the much stronger claim that without language there are no beliefs of which the content could be known. Stich argues that even if we could list all the objects that fit into a category somewhat similar to our bone category it would do us little good. We could get a list of objects which fit a particular list of properties, \( P_1 \& P_2 \& \ldots \& P_n \). But, Stich claims, even with this list we could be no closer to being able to attribute a belief to the dog:

...suppose there is a bone directly in front of Fido, clearly in view. Suppose further that it is one of those bones which is \( P_1 \& P_2 \& \ldots \& P_n \). Fido clearly recognizes it; he tenses with anticipation. What is it that he believes? We have already argued that we cannot capture the content of his belief by saying Fido believes that there is a bone in front of him.

And surely it is less plausible to suggest Fido believes there is an object in front of him which is \( P_1 \& P_2 \& \ldots \& P_n \). 'Generations of work by animal psychologists' may uncover a great deal about animal concepts. But it will be of no help in specifying the content of animal beliefs. (Stich 1979, 24)

The fact that we could go through the world and identify which objects Fido would categorize and treat as bones does not, in itself, give us the ability to attribute beliefs to him. This is not because Fido is a dog. The same is true in the case of human beings. If all I did was point to certain objects and say 'That is a bone' (I acted in no other way and could answer no further questions about bones) this would not be very good evidence that I have beliefs about bones, even if I picked out every bone in the world. If all we did was define a category of objects that Fido picks out and attach the label “bone” to it, this does not allow us to attribute beliefs about bones to him any
more than indicating (to use an example well-worn by recent philosophy of mind) exactly which objects a vending machine will accept as American quarters will justify us in attributing beliefs about American quarters to it. What is needed is a reason to call that division of the world a division in terms of beliefs. According to Stich and Davidson, Armstrong’s animal psychologists could not provide that.

Of course, if all they did was identify independent categories of objects that animals treat in different ways, then the interconnections needed to develop a system of meaningful belief contents would be missing. But what is to stop those studying the behavior of animals from doing so in a way that will allow them to actually describe a set of interconnected beliefs with specifiable contents? What must be added to Armstrong’s proposal, then, is an explanation of a methodology that will allow those studying the mental states of an animal to determine what the content of those states are, and how they are related to each other. Once this is accomplished, we can begin to test and revise belief attributions to animals in a meaningful way.

What follows is an argument to the effect that, while there may be things that might slow animal psychologists down when they attempt to attribute beliefs with content to animals, there is nothing that will necessarily stop them from doing so, if they go about things in the right way. I begin by pointing out that there are at least prima facie reasons to believe that there is quite a bit of similarity between the content of animal and human beliefs. From this point I support the claim that even if we take into account the differences between the background knowledge of animals and humans and the amount of indeterminacy in the beliefs of animals, we still find
enough content to allow for a meaningful interpretation of the belief states of many animals. I will conclude with a rough sketch of the kind of methodology we might use if we wish to accurately attribute content to the beliefs of animals. I will rely on the work of Barwise and Perry, as it can be applied to this problem.

I have argued that, even if a dog and its master have beliefs with different content, as in the cases presented above, they may still predict each other’s actions. Still, there are reasons to think that there is overlap between their beliefs. Malcolm’s dog, like Fido and all dogs, must act in its environment in very complex ways in order to survive. It must be able to differentiate between objects in order to act appropriately toward them and it must produce the appropriate behavior when it comes in contact with them. It does not swim to the top of the tree, drink the dirt around it, or let the cat walk peacefully by. It can act appropriately in an infinite variety of situations involving types of objects that are familiar to it. It treats many of these things in ways that make sense to us and, in many instances, it treats objects

95Descartes famously argued that animals do not have thoughts. His principal argument against attributing thoughts to animals is found in Part Five of his Discourse on Method. There, he argues that the only sure sign we have that a being thinks is its ability to use language. Using a language, Descartes argues, requires that one be able to respond appropriately to any of an infinite number of possible utterances. According to Descartes, it would not be possible to construct a machine which could do this. The difficulty in constructing a language-using machine, according to Descartes, is that language use depends on the ability to respond meaningfully to an infinite variety of sentences. Descartes claims that: “...it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men can do.” (Descartes 45-46) It is, Descartes maintains, the ability to respond appropriately to an infinite number of situations that is the indicator of whether or not reason is present.

It would be possible however, Descartes speculates, to construct machines which would be indistinguishable from animals: “...If any such machines had the organs and outer shape of a monkey or of some other animals that lacks reason, we should have no means of knowing that they did not possess entirely the same nature of these animals.” (Descartes 45) What I have argued here is that animals do seem to possess the ability to respond appropriately to an infinite variety of situations. In this sense, while they do not, in most cases, have the ability to use language, animals seem to have much in common with human beings.
in ways similar to the ways that humans would treat them. It takes in information about its environment in many of the same ways we do: it sees, feels, hears, smells and tastes. Malcolm and his dog live out their day-to-day lives in a common environment. We have some idea of how dogs' senses compare to our own and we also have seen the types of environments in which they have spent time. 96

Considered in light of these hopefully uncontroversial statements, the methodology recommended by Bennett, and the determination that content can be described in terms of situation semantics, many claims of the type Davidson makes are unconvincing. Davidson claims that it is hard to make sense of questions like: “Does the dog believe that the cat went up the oldest tree on the block?” But, we can make sense of such questions. Using terminology and methods taken from Bennett, we can make and support many claims about what information a dog registers by observing what kinds of things it has been exposed to and how it processes information about those things. From what we know about dogs there is no reason to

---

96 This type of argument goes back at least to Hume, who argued that because the outward behavior of animals is analogous to our own, we are justified in concluding that the causes of that behavior are similar the causes of our behavior. As he put it:

“'Tis from the resemblance of the external actions of animals to those we ourselves perform, that we judge their internal likewise to resemble ours; and the same principle of reasoning, carried one step farther, will make us conclude that since our internal actions resemble each other, the causes, from which they are deriv'd, must also be resembling.”

Hume continues by claiming that a common defect in theories of mind are that they describe thinking in ways according to which not only animals, but most people, could not be said to think: “The common defect of those systems, which philosophers have employ’d to account for the actions of the mind, is, that they suppose such a subtility and refinement of thought as not only exceeds the capacity of mere animals, but even of children and common people in our own species...” (David Hume, Treatise of Human Nature, ed. L.A. Selby-Bigge, (London: Oxford University Press, 1955) 176-7). The latter criticism of theories of mind may apply to contemporary theories of mind as much as it did to the theories of Hume’s time.
attribute the ability to pick out the oldest tree on the block as the oldest tree on the block to a dog. However, we might want to claim that the dog registers that this is the tree under which it recently buried a bone. There is no apparent reason why we cannot, on the basis of what we know about the way the dog senses and what is relevant to its general success in the world, decide which registrations the dog most likely has. We have good evidence that Fido may register that “the person who fills the food bowl,” is home. At the same time we can also be fairly secure in claiming that Fido does not register that the author of “Do Animals Have Beliefs?” is home. Of course, since Fido does not use language, Fido could not accept or reject these descriptions, but they may fit descriptions that would match objects that Fido would pick out, if they were presented in a way that allowed Fido to register these characteristics. For instance, showing Fido a copy of “Do Animals Have Beliefs?” would most likely not elicit any response that would show that Fido recognized that the author and Stich were the same person, but one could imagine Fido’s master walking in the door and Fido running to the cupboard in which his food is kept in many different situations, exhibiting behaviors that indicate that Fido recognizes that his master is the source of his food.

It does make sense to make such claims about Fido, and once we have begun doing so, filling in a more complete network of related content looks promising. The fact that Fido reacts appropriately to many objects is an indication that he possesses quite a bit of background knowledge. In order to make sense of much of Fido’s
behavior, we must assume that these pieces of knowledge are interconnected in various ways.

This is related to the point that, if we do begin to attribute beliefs about bones to Fido, then a key factor in those beliefs being beliefs about bones is that bones cause those beliefs. If it turns out to be the case that Fido treats a category of objects that is similar to the category of objects his master calls ‘bones’, in a way that is different from the way he treats other objects, then the master’s bone beliefs and Fido’s bone beliefs are likely caused by objects of the same type. For those with externalist tendencies, this will be significant in deciding the content of both Fido’s beliefs and mine. As I argued in Chapter Three,97 the recognition that ideas concerning the plausibility of explanations of how a believer came to have a particular belief help us decide between competing belief/desire explanations of behavior. In addition, as explained above, Barwise and Perry’s system allows us to indicate when an animal and a human have beliefs about the same object, but conceptualize it differently.

One who holds that animals do not have beliefs could maintain a consistent position by rejecting the claim that we can attribute meaningful registrations to animals, but, a methodology such as Bennett’s provides a way, in principle, at least, that gives reasons to think that we should be able to refine our attributions of interconnected networks of registrations, goals, and beliefs to animals. This is what Davidson claims, in his presentation of the specificity argument, that it is impossible to do. It might turn out that a research project launched along the lines that Bennett

---

97 See Chapter Three, 92-94.

217
advocates will not be productive, either because it is not practically feasible or because Fido does not have a network of interconnected goals and registrations. But, the problems in attributing beliefs to non-language-using animals are not the problems Davidson cites in his version of the specificity argument. It is not that we cannot make sense of any questions about the content of the registrations or beliefs of non-language-using animals. It is not clear what we will find, but we can explore the contents of such states and we can represent them using situation semantics.

The problem is even more serious for Stich’s position. Stich claims that while animals may be in states that have functional roles similar to the functional roles of human beliefs, because the content of any particular belief depends on background knowledge that non-language-using animals could not have, animals do not have beliefs. What Bennett’s work provides is a way of establishing what the content of an animal’s goals and registrations are by varying the situations that animal is in in such a way that the relationship between different registrations and goals is tested. So, while it may be that, in fact, no animals possess such a network of beliefs, if any did, we would have a way of finding out what the content of the beliefs in that network were, even if those beliefs were very different from our own. Attributing content to registrations is something we can do. The question is whether or not we will find that some animals have beliefs, rather than registrations only.

Stich argues that any human as ignorant about bones as Fido is would surely not be describable as having beliefs about bones. I doubt that he is correct on this point. Consider the ignorance of many human beings when it comes to certain
animals, for instance, snakes.\footnote{Richard Routley makes this point in his article, “Alleged Problems in attributing Beliefs, and Intentionality, to Animals”, Inquiry 24 (1981): 385-417.} This, in general, does not prohibit them from having beliefs about snakes. Furthermore, any human who was as ignorant about bones as Stich takes Fido to be would still lack many of the abilities that Fido obviously has when it comes to distinguishing bones from other things in the world. Dogs have a lot of information that the average human does not have when it comes to identifying bones. Of course they are also lacking in some areas in which humans excel, but it is not obvious that the ways in which the dog is lacking prohibit it from having a concept of bone. As a dog, to whom identifying bones might be quite important for survival, Fido could very well be better at identifying bones, and have a more nuanced store of knowledge about the way they taste and smell, than Stich or myself.

If we take this into account, and attribute a set of bone beliefs to Fido, then it begins to look more and more like we may be justified in attributing a whole belief system to Fido.

Stich is most likely correct that Fido is lacking beliefs about the anatomical functions of bones. Beliefs of this sort, at least that bones are part of the skeletal system and help provide rigidity to animal bodies, are central to the human concept of bone. This is why it is extremely important to note that the point is not that there must be an exact match between the content of Fido’s beliefs and the content of Stich’s. There are many things which are very important for human beings and of which Fido could never be aware. What there must be is \textit{enough of an overlap} in content to provide a \textit{starting point for interpretation}. This means only that animals must be
similar to humans to the degree that a methodology such as Bennett describes will work to refine our goal/registration theories in a way that helps us explain behavior. In reference to Malcolm’s dog, Fido, and other similar creatures, it seems that we are justified in assuming that this is true.

A very significant difference between animal and human beliefs may be that there is, in many cases, more indeterminacy in the content of the beliefs of animals. One reason for thinking this is because, as noted above, language allows human beings to make fine distinctions easily. Like the differences in the background knowledge of animals and humans, this difference should not be ignored, but it may seem less significant if we also examine human beliefs in this respect. In the case of human beings there is a great deal of variation between the belief structures and concepts of different individuals. I am sure that my beliefs about bones, when compared to those of an orthopedic surgeon, are lacking numerous connections to other beliefs and background knowledge and are therefore relatively indeterminate. My ignorance does not prevent me from having many beliefs about bones or even beliefs about the type of surgery required to repair a misalignment of the tibia and femur. Whether individual humans have fully determinate concepts is questionable. If there is greater indeterminacy in animal concepts than human concepts, this may cause difficulties in interpretation, but does not necessarily lead to an obvious rejection of the possibility that animals have beliefs. When it comes to animals, the question is whether or not the indeterminacy in the content of their concepts is so great that we cannot meaningfully attribute content to their concepts. This may not
completely answer Stich’s argument, however. It still needs to be shown how we may
describe the content of beliefs when that content differs significantly from the content
of human beliefs.

Few organisms, possibly only the higher mammals, seem like they
may possess complex sets of interconnected mental states. But, as argued in
Chapter Two, any organism that is highly educable and inquisitive does
possess such a set. So, Bennett provides us with a way of determining that a
non-language-using animal has such a set of beliefs and a methodology for
determining what the contents of those beliefs are.

According to the methodology I described in Chapter Two, we begin with a
hypothesis about a particular animal’s registration abilities and goals, then work at
testing and refining our attributions of registrations and goals by observing the animal
in various situations. We may, on the basis of these observations, determine that, even
if our attributions to the animal are generally correct, the animal does not have
beliefs, but only goals and registrations. Or, it may be that the animal is highly
educable about many propositions, including ones like the one under consideration,
and the animal is highly inquisitive about many propositions, including all-important
types to which the one in question belongs. If we find that this is the case, we attribute
beliefs to the animal. Bennett’s work provides us with a way of testing and refining
attributions of contentful registrations, goals, beliefs, and desires to non-language-
users.
Given the methodology Bennett proposes and the ability to represent content non-linguistically, the way Bennett has characterized the distinction between registrations and beliefs is a more useful distinction than Malcolm's thinking and having thoughts or Stich's beliefs and pseudo-beliefs, if what we are doing is attempting to find what capacities a being must have in order to have a thought. If the work of Bennett, as I presented it in Chapter Two, is accepted, and we can attribute content to the registrations and goals of animals, and we have a means of determining whether our attributions of content are correct, then it is not the fact that a state has highly specifiable content that determines whether or not it is a belief.

There may, for example, be a machine that registers very specific information, and has states that are inter-related in intricate ways. Still, if it is neither educable nor inquisitive, it would be problematic to call the states of the machine beliefs. This is true, as well, for any organism, whether it expresses the content of those states in sentences of a natural language or not. On the other end of the spectrum, there are many human beliefs that lack specific content — for example, my beliefs about bones or the theory of relativity or many other matters that I know little about — that should clearly be classified as beliefs. The key is not how specific the content of a state is, but how it fits into a larger spectrum of behaviors.

It does not follow from this that animals have beliefs, but it is clear that anyone who allows that contentful registrations can be meaningfully attributed to animals cannot reasonably claim that we know that animals do not have beliefs because it is impossible to accurately attribute content to their mental states. If
Bennett is correct in the claim that we can attribute content to registrations, then it is wrong to claim that if an animal does not have language, there is no way to explore the content of its mental states to the extent that such content may exist.

I have argued here that the adoption of a non-linguistic theory of content provides a way of accurately representing the beliefs of non-language-users. I argued further that Bennett's work helps show how it is that we can check and improve the accuracy of our attributions of content to the beliefs of animals. Still, if we are truly concerned about the accuracy of our attributions to Fido, or each other, there are some lessons to be learned from the specificity argument.

Perhaps most importantly, beliefs are not the kind of things that can be attributed one at a time. If Fido's mind is furnished with beliefs, as is frequently pointed out, they could not have been purchased individually; they are only available in large, inseparable sets. As Davidson explains, the content of any particular belief is dependent on that belief being located in a large network of other beliefs. In addition, as Stich argues, individual beliefs depend on the believer having relevant background knowledge. This is why establishing a category of things that Fido reacts to in a particular way, even if such a category matches up with our "bone" category, will not, on its own, allow us to attribute beliefs about bones to Fido. Instead, if we plan to attribute beliefs to Fido, our method will have to be roughly as follows: we
must begin with a few not insignificant assumptions about the way Fido breaks down and is motivated to act in the world.

In particular we must assume, at least tentatively, that Fido is rational, that Fido has registrations and goals, and that Fido sees the world in terms of objects that are meaningful to him. In short, we assume that Fido is an intentional being or adopt something like the intentional stance toward him. On the basis of these assumptions and some general knowledge about Fido we can begin to attribute a network of beliefs to him. We can expect our initial network to have many problems. These will include wrongly attributed and indeterminate content as well as mistaken ideas about how Fido’s registrations and goals are interconnected.

From this point, following Bennett, we study Fido in various situations, and dealing with various objects, humans and other animals. We can then fine-tune our conception of Fido’s network of registrations and goals until we are content with our level of accuracy. We are then in a position to ask whether or not the animal has beliefs. We do this by determining whether or not the animal is highly educable about a number of propositions similar to the ones about which we are inquiring and by determining whether or not the animal is inquisitive about such propositions as well. This process is helped by the fact that the acceptance of a non-linguistic theory of content eliminates problems stemming from the fact that animals possess different background

---

99This is not to say that Dennett’s intentional stance theory will, in the end, provide the best explanation of belief, it is only to say that something like this stance must be adopted if we are to begin interpretation.
beliefs from those of humans, and that it may be very difficult to capture the
ccontent of those beliefs in sentences of natural language.

I have attempted to show how Bennett’s work establishes that it is
possible to attribute to non-language-users the knowledge and background
beliefs that are necessary in order to construct an initial network of beliefs.
This framework also allows us to attribute further beliefs with which we can
fine-tune that system. I have also argued that Stalnaker’s work lays a
foundation for representing the intentional states of non-language-users in a
way that accurately reflects the interrelations between such states. This theory
can then be built on by incorporating the work of Barwise and Perry. The
problems expressed by Stich and Davidson indicate considerable difficulties
on the practical side of attributing beliefs to non-human animals. I have
argued that despite these difficulties, it is possible in many cases, through the
adoption of a non-linguistic account of content, to meaningfully and
accurately attribute intentional states to animals and to determine whether or
not it makes sense to call such states beliefs.
Works Cited


